

UNIVERSIDAD PABLO DE OLAVIDE
DEPARTAMENTO DE ECONOMÍA, MÉTODOS CUANTITATIVOS E
HISTORIA ECONÓMICA



Segmentación de usuarios en la oficina de farmacia mediante algoritmos bioinspirados

**TESIS DOCTORAL DEL PROGRAMA DE DOCTORADO EN
ESTADÍSTICA E INVESTIGACIÓN OPERATIVA**

Manuel Martínez Blanes

Sevilla, noviembre de 2015

UNIVERSIDAD PABLO DE OLAVIDE
DEPARTAMENTO DE ECONOMÍA, MÉTODOS CUANTITATIVOS E
HISTORIA ECONÓMICA



Memoria presentada por Manuel Martínez Blanes para optar al Grado de Doctor por la
Universidad Pablo de Olavide de Sevilla

Fdo: Manuel Martínez Blanes

Vº Bº

Los Directores de la Tesis Doctoral

Fdo.: Dr. Alfredo García Hernández-Díaz

Departamento de Economía, Métodos
Cuantitativos e Historia Económica

Universidad Pablo de Olavide

Fdo.: Dr. Antonio M^a. Rabasco Álvarez

Departamento de Farmacia y Tecnología
Farmacéutica

Universidad de Sevilla

Sevilla, noviembre de 2015

A mis tres soles...

AGRADECIMIENTOS

En primer lugar, mi agradecimiento, reconocimiento y respecto a mis directores de tesis Alfredo García y Antonio M^a Rabasco. Ambos han sabido conjugar su experiencia y su capacidad de trabajo y rigurosidad. Hecho que me ha servido de acicate para no cejar en el empeño de llevar a buen puerto este barco. Gracias de corazón.

A Manuel Ojeda, farmacéutico y soñador a partes iguales. Con él he compartido este camino. Hemos charlado eternamente buscando nuevas formas, nuevos retos que alcanzar desde la oficina de farmacia, eso sí, bajo la máxima “en el mostrador tiene que funcionar...”. Agracedimiento que hago extensivo a todo el personal de su farmacia.

A la cátedra Avenzoar, por su apoyo, acogida y soporte en este trabajo.

A Manuel Pérez, en nombre del Colegio de Farmacéuticos de Sevilla, por abrirme las puertas de la organización colegial y poder pulsar la realidad de las farmacias de la provincia de Sevilla en lo que respecta al tema central de la investigación.

A Concha Almarza, un ejemplo de que si hay buenas personas detrás de las redes sociales, éstas son útiles.

A Antonio Mingorance y Francisco Peinado, en nombre del Consejo Andaluz de Colegios Oficiales Farmacéuticos, por facilitarme el acceso a información útil para este trabajo y compartir mis inquietudes.

A Antonio Peinado y Alicia Aguilar, profesionales incansables. Con ellos descubrí el mundo farmacéutico y muchas más cosas...todas buenas.

A mi amigo Paco Pérez Torres, por las horas y horas de clases maestras: sanidad, relaciones humanas, grandes proyectos...

A Luciano Barrios, amigo y médico de atención primaria, que me enseñó que “el rigor lo pone el profesional”.

A mi gran mentor y amigo José Manuel Velarde, un físico atípico de calidad humana infinita que siempre ha estado ahí. En esta aventura también.

A Carlos Marchante, por permitirme la confrontación de los puntos de vista de la ingeniería y la estadística.

A toda mi familia, a mis amigas y amigos y siempre a ti Fernando Ortega...

A Loli y Manolo, mis queridos padres, me lo han dado todo sin pedir nada a cambio. En especial que, tanto mi hermana Isabel como yo, fuésemos buenas personas.

A Fernando, mi príncipe de caramelo y a Celia, mi princesa de chicle, dos de mis soles. Vosotros me dais la fuerza y me hacéis soñar con el futuro.

A ti Toñi, mi tercer sol, por dejarte robar tantas y tantas horas y fines de semana. Por ponérmelo tan fácil, por tu apoyo infinito, tu ternura y tu amor.

A todas las personas, muchas, que me han permitido empaparme de ideas, conceptos, experiencias que he hecho mías y que deambulan por este trabajo. Por supuesto, mi agradecimiento a aquellos que en el algún momento me dijeron “no” y crearon en mí el afán de superación.

Gracias a todos...

ÍNDICE

Índice

| | | |
|------------|--|------------|
| 1 | INTRODUCCIÓN Y ENTENDIMIENTO DEL PROBLEMA | 1 |
| 1.1 | OBJETIVOS | 6 |
| 1.2 | EL ENTORNO DE LA OFICINA DE FARMACIA | 8 |
| 1.2.1 | LOS MODELOS FARMACÉUTICOS | 9 |
| 1.2.1.1 | Los modelos farmacéuticos de Estados Unidos y Japón | 14 |
| 1.2.1.2 | Los modelos farmacéuticos europeos | 18 |
| 1.2.1.3 | El modelo farmacéutico español | 29 |
| 1.2.2 | LOS CANALES DE DISTRIBUCIÓN Y DISPENSACIÓN DE MEDICAMENTOS EN ESPAÑA | 36 |
| 1.2.3 | LA OFICINA DE FARMACIA ESPAÑOLA EN CIFRAS | 43 |
| 1.2.4 | EL GASTO SANITARIO EN ESPAÑA | 50 |
| 1.2.5 | EL MERCADO FARMACÉUTICO ESPAÑOL | 54 |
| 1.2.5.1 | El mercado de los medicamentos de prescripción médica | 55 |
| 1.2.5.2 | Mercado del autocuidado de la salud | 90 |
| 1.2.6 | SITUACIÓN DE LAS OFICINAS DE FARMACIA | 100 |
| 1.2.7 | RETOS Y TENDENCIAS DE FUTURO ANALIZADOS | 113 |
| 1.3 | HERRAMIENTAS PARA EL DESARROLLO DE LA OFICINA DE FARMACIA | 115 |
| 1.3.1 | PRODUCTOS Y SERVICIOS DE LA OFICINA DE FARMACIA | 117 |
| 1.3.2 | PACIENTES Y CLIENTES: USUARIOS | 122 |
| 1.3.3 | MARKETING FARMACÉUTICO | 126 |
| 1.3.3.1 | Fidelización de clientes | 129 |
| 1.3.3.2 | Registro de pacientes y clientes | 143 |
| 1.3.3.3 | Segmentación de clientes | 145 |

| | | |
|------------|--|------------|
| 1.4 | EL PROBLEMA ESTADÍSTICO DE LA SEGMENTACIÓN | 154 |
| 1.4.1 | INTRODUCCIÓN | 154 |
| 1.4.2 | SEGMENTACIÓN FRENTE A CLASIFICACIÓN | 157 |
| 1.4.3 | SEGMENTACIÓN DE BASES DE DATOS. PROS Y CONTRAS | 159 |
| 1.4.4 | DETERMINACIÓN DEL NÚMERO DE CLÚSTERES | 162 |
| 1.4.4.1 | Regla general | 163 |
| 1.4.4.2 | El método del codo (<i>Elbow method</i>) | 164 |
| 1.4.4.3 | Indicadores | 165 |
| 1.4.4.4 | Segmentación espectral | 167 |
| 1.4.4.5 | Algoritmos evolutivos | 168 |
| 1.4.5 | INDICADORES DE VALIDEZ | 168 |
| 1.4.5.1 | Indicador de Calinski-Harabasz | 171 |
| 1.4.5.2 | Indicador de Davies-Bouldin | 172 |
| 1.4.5.3 | Indicador de Dunn | 173 |
| 1.4.5.4 | Variantes del indicador de Dunn | 174 |
| 1.4.5.5 | Indicador CS | 175 |
| 1.4.6 | SEGMENTACIÓN MULTIOBJETIVO | 175 |
| 1.4.7 | ALGORITMOS DE SEGMENTACIÓN | 178 |
| 1.4.7.1 | Algoritmos de particionamiento: <i>K-means</i> y variantes | 184 |
| 1.4.7.2 | Algoritmos jerárquicos: aglomerativos y divisivos | 192 |
| 1.4.7.3 | Algoritmos metaheurísticos | 199 |
| 2 | METODOLOGÍA | 235 |

| | | |
|------------|--|------------|
| 2.1 | EL ALGORITMO DE EVOLUCIÓN DIFERENCIAL PARA SEGMENTACIÓN CON ESQUEMA DE CONTINUACIÓN | 237 |
| | (DECCS) | 237 |
| 2.1.1 | INTRODUCCIÓN | 237 |
| 2.1.2 | CODIFICACIÓN | 239 |
| 2.1.3 | EVALUACIÓN DE LA CALIDAD. FUNCIÓN OBJETIVO | 241 |
| 2.1.4 | POBLACIÓN INICIAL | 242 |
| 2.1.5 | MUTACIÓN Y CRUCE | 243 |
| 2.1.6 | EL ESQUEMA DE CONTINUACIÓN | 244 |
| 2.1.6.1 | Continuación. Falta de diversidad en los vectores diferencia | 244 |
| 2.1.6.2 | Perturbaciones grandes | 246 |
| 2.1.7 | PSEUDOCÓDIGO DE DECCS | 248 |
| 2.2 | BENCHMARK EN SEGMENTACIÓN | 250 |
| 2.2.1 | PROBLEMAS DE TAMAÑO MEDIO | 252 |
| 2.2.2 | PROBLEMAS DE GRAN TAMAÑO | 255 |
| 2.2.3 | CONJUNTO DE DATOS DE UNA OFICINA DE FARMACIA | 258 |
| 2.3 | SOFTWARE Y HARDWARE UTILIZADOS | 259 |
| 2.4 | ENCUESTA A FARMACÉUTICOS DE LA PROVINCIA DE SEVILLA | 260 |
| 3 | RESULTADOS Y DISCUSIÓN | 263 |
| 3.1 | INTRODUCCIÓN | 265 |
| 3.2 | LA FUNCIÓN OBJETIVO. INDICADORES DE CALIDAD | 266 |
| 3.3 | SIMETRÍA EN LA CODIFICACIÓN | 276 |
| 3.4 | COMPARATIVA DE MÉTODOS DE MUESTREO PARA LA POBLACIÓN INICIAL | 285 |
| 3.5 | COMPARATIVA DE DECCS FRENTE A OTROS ALGORITMOS DE SEGMENTACIÓN | 301 |

| | | |
|----------|--|------------|
| 3.5.1 | COMPARATIVA DE RESULTADOS | 302 |
| 3.5.2 | VELOCIDADES Y TIEMPOS DE CONVERGENCIA DE DECCS | 312 |
| 3.5.3 | RESULTADOS EN BASES DE DATOS MUY GRANDES | 313 |
| 3.5.4 | APLICACIÓN A LA OFICINA DE FARMACIA | 315 |
| 3.6 | RESULTADOS DE LA ENCUESTA A FARMACÉUTICOS DE LA PROVINCIA DE SEVILLA | 320 |
| 4 | RESUMEN Y CONCLUSIONES | 331 |
| 4.1 | PRINCIPALES APORTACIONES | 333 |
| 4.2 | CONCLUSIONES | 335 |
| 4.3 | LIMITACIONES DE LA INVESTIGACIÓN | 340 |
| 4.4 | TRABAJOS FUTUROS Y NUEVAS LÍNEAS DE INVESTIGACIÓN | 341 |
| 5 | BIBLIOGRAFÍA | 343 |
| 6 | ANEXOS | 371 |
| 6.1 | CUESTIONARIO TARJETA FIDELIZACIÓN | 373 |
| 6.2 | MUESTREOS REALIZADOS PARA 20, 30, 40 Y 100 SOLUCIONES | 374 |
| 6.2.1 | PROBLEMA IRIS | 374 |
| 6.2.2 | PROBLEMA GLASS | 376 |
| 6.2.3 | PROBLEMA S1 | 378 |
| 6.2.4 | PROBLEMA YEAST | 380 |
| 6.3 | CUESTIONARIO A FARMACÉUTICOS DE SEVILLA | 382 |
| 6.4 | CÓDIGO C++ DEL ALGORITMO DECCS UTILIZADO EN LA INVESTIGACIÓN | 383 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| FIGURA 1. CADENA DE VALOR DEL SECTOR SANITARIO | 8 |
| FIGURA 2. MODELOS DE FARMACIA..... | 9 |
| FIGURA 3. PORCENTAJE DE ESTADOS DE LA UNIÓN EUROPEA DE LOS 27 QUE PRESENTAN LA CARACTERÍSTICA MEDIDA EN SU SISTEMA FARMACÉUTICO..... | 25 |
| FIGURA 4. CARACTERÍSTICAS DEL MODELO FARMACÉUTICO ESPAÑOL..... | 30 |
| FIGURA 5. MAPA ESPAÑOL DE COBERTURA DE OFICINAS DE FARMACIA | 34 |
| FIGURA 6. CANALES DE DISTRIBUCIÓN Y DISPENSACIÓN. CUOTAS Y MÁRGENES..... | 37 |
| FIGURA 7. EVOLUCIÓN DEL NÚMERO DE FARMACIAS Y DEL RATIO HABITANTES/FARMACIA EN ESPAÑA (2005-2013) | 44 |
| FIGURA 8. EVOLUCIÓN DE LA POBLACIÓN ESPAÑOLA SEGÚN PADRÓN EN EL PERÍODO (2003-2013) | 45 |
| FIGURA 9. EVOLUCIÓN DEL NÚMERO DE COLEGIADOS Y TITULARES DE FARMACIA EN ESPAÑA EN EL PERÍODO (2005-2013)..... | 48 |
| FIGURA 10. EVOLUCIÓN DEL NÚMERO DE COLEGIADOS SEGÚN SEXO EN ESPAÑA EN EL PERÍODO (2005- 2013)..... | 48 |
| FIGURA 11. EVOLUCIÓN DEL NÚMERO DE TITULARES SEGÚN SEXO EN ESPAÑA EN EL PERÍODO (2005- 2013)..... | 49 |
| FIGURA 12. EVOLUCIÓN DEL RATIO TITULARES/COLEGIADOS EN ESPAÑA EN EL PERÍODO (2005-2013) | 50 |
| FIGURA 13. CANALES, TIPOS DE MEDICAMENTOS Y SUS CUOTAS EN EL MERCADO FARMACÉUTICO SEGÚN PRECIO..... | 54 |
| FIGURA 14. EL MERCADO DEL MEDICAMENTO Y SUS PRINCIPALES AGENTES | 56 |
| FIGURA 15. CUOTA DE MERCADO DE MEDICAMENTOS A PRECIO DE GENÉRICO | 62 |
| FIGURA 16. TASA DE VARIACIÓN INTERANUAL (DICIEMBRE 2014 RESPECTO DICIEMBRE 2013) | 65 |

| | |
|---|----|
| FIGURA 17. AHORRO PROVOCADO POR EL USO DE GENÉRICOS (MILLONES DE €). AÑO 2010..... | 66 |
| FIGURA 18. ACCESO DE MEDICAMENTOS, PRODUCTOS MÉDICOS Y DE DIAGNÓSTICO AL HOSPITAL | 67 |
| FIGURA 19. MERCADO HOSPITALARIO ESPAÑOL SEGÚN IMS HEALTH | 71 |
| FIGURA 20. MERCADO HOSPITALARIO ESPAÑOL SEGÚN FARMAINDUSTRIA | 72 |
| FIGURA 21. EVOLUCIÓN DEL NÚMERO DE RECETAS Y GASTO FARMACÉUTICO POR RECETAS EN EL PERIODO 2008-2014..... | 76 |
| FIGURA 22. EVOLUCIÓN DEL GASTO FARMACÉUTICO POR HABITANTE EN EL PERIODO 2008-2014 | 76 |
| FIGURA 23. EVOLUCIÓN DEL GASTO FARMACÉUTICO POR RECETAS Y EL GASTO MEDIO POR RECETA EN EL PERIODO 2008-2014 | 77 |
| FIGURA 24. TVMA DEL NÚMERO DE RECETAS FACTURADAS DEL SISTEMA NACIONAL DE SALUD EN EL PERIODO 2008-2014 | 79 |
| FIGURA 25. NÚMEROS ÍNDICE, BASE=2008, DEL NÚMERO DE RECETAS FACTURADAS DEL SISTEMA NACIONAL DE SALUD | 80 |
| FIGURA 26. CUOTAS DE PARTICIPACIÓN SOBRE EL TOTAL DE RECETAS FACTURADAS Y POBLACIÓN (INE 2014)..... | 81 |
| FIGURA 27. TVMA DEL GASTO FARMACÉUTICO A TRAVÉS DE RECETA OFICIAL DEL SISTEMA NACIONAL DE SALUD EN EL PERIODO 2008-2014..... | 84 |
| FIGURA 28. NÚMEROS ÍNDICE, BASE=2008, DEL GASTO FARMACÉUTICO A TRAVÉS DE RECETA OFICIAL DEL SISTEMA NACIONAL DE SALUD..... | 85 |
| FIGURA 29. CUOTAS DE PARTICIPACIÓN SOBRE EL TOTAL DEL GASTO FARMACÉUTICO EN RECETAS Y POBLACIÓN (INE 2014) | 86 |

| | |
|---|-----|
| FIGURA 30. EVOLUCIÓN DEL GASTO MEDIO POR RECETA FACTURADA DEL SISTEMA NACIONAL DE SALUD EN EL PERIODO 2008-2014 | 88 |
| FIGURA 31. COMPARACIÓN DEL GASTO MEDIO POR RECETA POR COMUNIDAD AUTÓNOMA | 89 |
| FIGURA 32. SEGMENTOS EN EL MERCADO DEL AUTOCUIDADO | 90 |
| FIGURA 33. DISTRIBUCIÓN DE LOS SEGMENTOS DEL MERCADO DE CONSUMER HEALTH (%€ PVP; 2014) | 92 |
| FIGURA 34. DISTRIBUCIÓN DE LAS SEGMENTOS Y CATEGORÍAS DEL MERCADO DE CONSUMER HEALTH (%€ PVP; 2014)..... | 93 |
| FIGURA 35. EVOLUCIÓN DEL MERCADO DE AUTOCUIDADO EN VALORES (PVP €) Y UNIDADES. 2010-2014 | 94 |
| FIGURA 36. EVOLUCIÓN DEL MERCADO DE AUTOCUIDADOS, SEGÚN SEGMENTOS, EN VALORES (PVP €). 2010-2014..... | 96 |
| FIGURA 37. EVOLUCIÓN DEL MERCADO DE AUTOCUIDADOS, SEGÚN SEGMENTOS, EN MILLONES DE UNIDADES. 2010-2014 | 97 |
| FIGURA 38. PANTALLA AZUL CON LA QUE DEBE FINALIZAR CUALQUIER PUBLICIDAD DE MEDICAMENTOS PUBLICITARIOS EN MEDIOS AUDIOVISUALES | 99 |
| FIGURA 39. EVOLUCIÓN DEL MARGEN NETO DE UNA OFICINA DE FARMACIA TIPO (*) (1996-2012) . | 100 |
| FIGURA 40. LA SITUACIÓN DE LOS IMPAGOS EN LAS COMUNIDADES AUTÓNOMAS.2014 | 102 |
| FIGURA 41. EVOLUCIÓN DEL MERCADO DE PRESCRIPCIÓN Y CAMBIOS REGULATORIOS 2007-2014.... | 104 |
| FIGURA 42. VALORACIÓN DE LA FARMACIA HABITUAL EN LOS AÑOS 2001 Y 2009 | 108 |
| FIGURA 43. MATRIZ DAFO DE LA OFICINA DE FARMACIA ESPAÑOLA | 112 |

| | |
|--|-----|
| FIGURA 44. FACTORES QUE INCIDIRÁN EN EL MERCADO FARMACÉUTICO ESPAÑOL DE FORMA POSITIVA O NEGATIVA..... | 113 |
| FIGURA 45. ZONA DE CONFORT Y ZONA DE APRENDIZAJE | 117 |
| FIGURA 46. CLASIFICACIÓN DE LOS PRODUCTOS EN UNA OFICINA DE FARMACIA..... | 118 |
| FIGURA 47. OFICINA DE FARMACIA Y OTROS ESTABLECIMIENTOS SIMILARES | 119 |
| FIGURA 48. SERVICIOS OFRECIDOS EN UNA OFICINA DE FARMACIA | 121 |
| FIGURA 49. TIPOS DE USUARIOS EN UNA OFICINA DE FARMACIA | 124 |
| FIGURA 50. TIPOLOGÍA DE VENTAS Y USUARIOS EN UNA OFICINA DE FARMACIA | 125 |
| FIGURA 51. RELACIÓN ENTRE LA VIDA MEDIA Y LA TASA DE RETENCIÓN DE CLIENTES | 132 |
| FIGURA 52. ECUACIÓN DE VALOR PARA EL CLIENTE | 133 |
| FIGURA 53. BIG DATA | 155 |
| FIGURA 54. DIFERENCIAS ENTRE SEGMENTACIÓN Y CLASIFICACIÓN | 158 |
| FIGURA 55. NUBE DE PUNTOS EN DOS DIMENSIONES | 159 |
| FIGURA 56. EJEMPLO DE TRES CLÚSTER..... | 160 |
| FIGURA 57. DIFERENTES GEOMETRÍAS DE CONGLOMERADOS..... | 162 |
| FIGURA 58. MÉTODO DE ELBOW | 165 |
| FIGURA 59. EJEMPLO DE LA EVOLUCIÓN DEL INDICADOR DE CALINSKY-HARABASZ | 166 |
| FIGURA 60. GRAFO DE PROXIMIDAD O SIMILITUD | 167 |
| FIGURA 61. DISTINTAS GEOMETRÍAS QUE PRESENTAN LAS CIRCUNFERENCIAS DE RADIO SEGÚN VALORES DE P..... | 170 |
| FIGURA 62. GEOMETRÍA USANDO LA DISTANCIA DE MAHALANOBIS | 170 |
| FIGURA 63. DISTANCIA COSENO ENTRE DOS PUNTOS | 171 |

| | |
|--|-----|
| FIGURA 64. PARTICIONADO MULTIOBJETIVO | 176 |
| FIGURA 65. SOLUCIONES PARETO Y FRENTE DE PARETO | 177 |
| FIGURA 66. LA FORMACIÓN DE CLÚSTERES..... | 178 |
| FIGURA 67. AGRUPAMIENTO DIFUSO DE DOS GRUPOS. LAS ZONAS OSCURAS INDICAN ALTOS GRADOS DE PERTENENCIA. BASADO EN SOSA Y COL. (159)..... | 181 |
| FIGURA 68. CLASIFICACIÓN DE LOS ALGORITMOS DE SEGMENTACIÓN | 182 |
| FIGURA 69. EJEMPLO DENDROGRAMA | 192 |
| FIGURA 70. DISTANCIAS ENTRE CLÚSTERES: A) VECINO MÁS CERCANO B) MÁS LEJANO C) PROMEDIO D) CENTROIDE O CENTRO DE GRAVEDAD..... | 194 |
| FIGURA 71. CLASIFICACIÓN DE ALGORITMOS DE CLASIFICACIÓN (ADAPTADO Y AMPLIADO DE (134)) .. | 201 |
| FIGURA 72. CLASIFICACIÓN DE LOS ALGORITMOS DE COMPUTACIÓN EVOLUTIVA | 202 |
| FIGURA 73. DIAGRAMA DE FLUJO DE UN GA..... | 212 |
| FIGURA 74. REPRESENTACIÓN CONCEPTUAL DE LAS EES Y LOS AGs. | 216 |
| FIGURA 75. OPERACIONES DE CRUCE Y MUTACIÓN EN PROGRAMACIÓN GENÉTICA | 218 |
| FIGURA 76. ETAPAS DEL ALGORITMO DE..... | 223 |
| FIGURA 77. ALGORITMO DE..... | 228 |
| FIGURA 78. MUESTRA DE TAMAÑO 4 MEDIANTE HIPERCUBOS LATINOS PARA UN ESPACIO BIDIMENSIONAL | 243 |
| FIGURA 79. EJEMPLO DE FUNCIÓN CDF(x) | 245 |
| FIGURA 80. IRIS | 252 |
| FIGURA 81. GLASS | 253 |
| FIGURA 82. WINE | 253 |

| | |
|---|-----|
| FIGURA 83. WINSCONSIN BREAST CANCER | 253 |
| FIGURA 84. THYROID | 254 |
| FIGURA 85. S-SET | 254 |
| FIGURA 86. LED DISPLAY DOMAIN DATA SET | 255 |
| FIGURA 87. YEAST | 256 |
| FIGURA 88. RED WINE QUALITY DATA SET..... | 257 |
| FIGURA 89. PEN-BASED RECOGNITION OF HANDWRITTEN DIGITS DATA SET | 257 |
| FIGURA 90. SONAR, MINES VS ROCKS DATA SET | 258 |
| FIGURA 91. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S1 TRUNCADO A 3 GRUPOS..... | 268 |
| FIGURA 92. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S1 TRUNCADO A 5 GRUPOS..... | 268 |
| FIGURA 93. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S1 TRUNCADO A 7 GRUPOS..... | 269 |
| FIGURA 94. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S2 TRUNCADO A 3 GRUPOS..... | 269 |
| FIGURA 95. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S2 TRUNCADO A 5 GRUPOS..... | 270 |
| FIGURA 96. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S2 TRUNCADO A 7 GRUPOS..... | 270 |
| FIGURA 97. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S3 TRUNCADO A 3 GRUPOS..... | 271 |
| FIGURA 98. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S3 TRUNCADO A 5 GRUPOS..... | 272 |
| FIGURA 99. CONGLOMERADOS ÓPTIMOS PARA EL PROBLEMA S3 TRUNCADO A 7 GRUPOS..... | 272 |
| FIGURA 100. SOLUCIÓN OBTENIDA POR DECCS PARA S1 TRUNCADO PARA 3 GRUPOS CON LA MÉTRICA CS | 274 |
| FIGURA 101. VELOCIDADES DE CONVERGENCIA MEDIA PARA IRIS | 278 |
| FIGURA 102. VELOCIDADES DE CONVERGENCIA MEDIA PARA GLASS | 279 |
| FIGURA 103. VELOCIDADES DE CONVERGENCIA MEDIA PARA WINE | 279 |

| | |
|--|-----|
| FIGURA 104. VELOCIDADES DE CONVERGENCIA MEDIA PARA BREAST | 280 |
| FIGURA 105. VELOCIDADES DE CONVERGENCIA MEDIA PARA THYROID..... | 280 |
| FIGURA 106. VELOCIDADES DE CONVERGENCIA MEDIA PARA LED | 282 |
| FIGURA 107. VELOCIDADES DE CONVERGENCIA MEDIA PARA YEAST | 283 |
| FIGURA 108. VELOCIDADES DE CONVERGENCIA MEDIA PARA WINE QUALITY | 283 |
| FIGURA 109. VELOCIDADES DE CONVERGENCIA MEDIA PARA SONAR..... | 284 |
| FIGURA 110. VECTORES OPUESTOS EN LA RECTA REAL..... | 287 |
| FIGURA 111. VECTORES OPUESTOS EN EL PLANO | 287 |
| FIGURA 112. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS ALEATORIOS PARA EL PROBLEMA IRIS | 289 |
| FIGURA 113. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS USANDO HIPERCUBOS LATINOS PARA EL PROBLEMA IRIS..... | 289 |
| FIGURA 114. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS USANDO VECTORES OPUESTOS PARA EL PROBLEMA IRIS | 290 |
| FIGURA 115. CALIDAD MEDIA PARA 51 MUESTREOS USANDO LOS TRES MÉTODOS PROPUESTOS PARA EL PROBLEMA IRIS CON TAMAÑOS DE POBLACIÓN 20, 30, 40 Y 100..... | 290 |
| FIGURA 116. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS ALEATORIOS PARA EL PROBLEMA GLASS | 292 |
| FIGURA 117. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS USANDO HIPERCUBOS LATINOS PARA EL PROBLEMA GLASS | 293 |
| FIGURA 118. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS USANDO VECTORES OPUESTOS PARA EL PROBLEMA GLASS..... | 293 |

| | |
|---|-----|
| FIGURA 119. CALIDAD MEDIA PARA 51 MUESTREOS USANDO LOS TRES MÉTODOS PROPUESTOS PARA EL PROBLEMA GLASS CON TAMAÑOS DE POBLACIÓN 20, 30, 40 Y 100 | 294 |
| FIGURA 120. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS ALEATORIOS PARA EL PROBLEMA S1 | 295 |
| FIGURA 121. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS USANDO HIPERCUBOS LATINOS PARA EL PROBLEMA S1 | 296 |
| FIGURA 122. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS USANDO VECTORES OPUESTOS PARA EL PROBLEMA S1 | 296 |
| FIGURA 123. CALIDAD MEDIA PARA 51 MUESTREOS USANDO LOS TRES MÉTODOS PROPUESTOS PARA EL PROBLEMA S1 CON TAMAÑOS DE POBLACIÓN 20, 30, 40 Y 100..... | 297 |
| FIGURA 124. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS ALEATORIOS PARA EL PROBLEMA YEAST | 298 |
| FIGURA 125. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS USANDO HIPERCUBOS LATINOS PARA EL PROBLEMA YEAST | 299 |
| FIGURA 126. CALIDAD MEDIA, MEJOR Y PEOR VALOR OBTENIDO PARA 51 MUESTREOS USANDO VECTORES OPUESTOS PARA EL PROBLEMA YEAST | 299 |
| FIGURA 127. CALIDAD MEDIA PARA 51 MUESTREOS USANDO LOS TRES MÉTODOS PROPUESTOS PARA EL PROBLEMA YEAST CON TAMAÑOS DE POBLACIÓN 20, 30, 40 Y 100..... | 300 |
| FIGURA 128. RESULTADOS COMPARATIVOS ENTRE DECCS Y ACDE EN IRIS..... | 303 |
| FIGURA 129. RESULTADOS COMPARATIVOS ENTRE DECCS Y ACDE EN GLASS..... | 304 |
| FIGURA 130. RESULTADOS COMPARATIVOS ENTRE DECCS Y ACDE EN WINE..... | 305 |
| FIGURA 131. RESULTADOS COMPARATIVOS ENTRE DECCS Y ACDE EN BREAST | 306 |

| | |
|--|-----|
| FIGURA 132. RESULTADOS COMPARATIVOS ENTRE DECCS Y ACDE EN THYROID | 307 |
| FIGURA 133. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN LED | 308 |
| FIGURA 134. RESULTADOS COMPARATIVOS ENTRE DECCS Y ACDE EN YEAST | 309 |
| FIGURA 135. RESULTADOS COMPARATIVOS ENTRE DECCS Y ACDE EN WINE QUALITY | 310 |
| FIGURA 136. RESULTADOS COMPARATIVOS ENTRE DECCS Y ACDE EN SONAR | 311 |
| FIGURA 137. COMPARATIVA DE LA VELOCIDAD MEDIA DE CONVERGENCIA ENTRE DECCS Y ACDE | 315 |
| FIGURA 138. COMPOSICIÓN DE LOS GRUPOS OBTENIDOS POR DECCS EN LA OFICINA DE FARMACIA SELECCIONADA | 316 |
| FIGURA 139. REPARTO DE LAS VENTAS ENTRE LAS CATEGORÍAS “VENTA LIBRE” Y “VENTA SOE” | 317 |
| FIGURA 140. TIPO DE VENTA SEGÚN CLÚSTERES | 318 |
| FIGURA 141. TIPO DE FARMACIA Y PRESENCIA ÚNICA O NO EN LA LOCALIDAD | 322 |
| FIGURA 142. TIPO DE FARMACIA SEGÚN SI LA DISTANCIA AL CENTRO DE SALUD MÁS PRÓXIMO ES INFERIOR A 250 METROS..... | 323 |
| FIGURA 143. COMPARACIÓN DE LA DISTINCIÓN ENTRE LA CONSIDERACIÓN DE “CLIENTES” Y “PACIENTES” Y SI LA FARMACIA POSEE O NO TARJETA DE FIDELIZACIÓN PARA FARMACIAS QUE NO SON ÚNICAS EN SU LOCALIDAD | 326 |
| FIGURA 144. COMPARACIÓN DE LA DISTINCIÓN ENTRE LA CONSIDERACIÓN DE “CLIENTES” Y “PACIENTES” Y SI LA FARMACIA POSEE O NO TARJETA DE FIDELIZACIÓN PARA FARMACIAS QUE SON ÚNICAS EN SU LOCALIDAD | 326 |
| FIGURA 145. MEDIO DE COMUNICACIÓN MÁS UTILIZADO | 328 |
| FIGURA 146. INFORMACIÓN REGISTRADA DE CLIENTES O PACIENTES | 330 |

ÍNDICE DE TABLAS

| | |
|---|----|
| TABLA 1. TIPOS DE SISTEMAS SANITARIOS SEGÚN FUNCIONES Y PAÍSES..... | 11 |
| TABLA 2. COPAGO ANTES Y DESPUÉS DEL RD 16/2012 | 13 |
| TABLA 3. ESTABLECIMIENTOS DE VENTA DE MEDICAMENTOS EN JAPÓN..... | 17 |
| TABLA 4. COMPARACIÓN DE LOS DIFERENTES MODELOS FARMACÉUTICOS EN LA UE27 | 20 |
| TABLA 5. COMPARATIVA DEL NÚMERO DE FARMACIAS Y POBLACIÓN EN LA UE27 | 22 |
| TABLA 6. COMPARATIVA DEL NÚMERO DE FARMACIAS Y HABITANTES POR FARMACIA SEGÚN LA PROPIEDAD FARMACÉUTICA | 23 |
| TABLA 7. COMPARATIVA DEL NÚMERO DE FARMACIAS Y HABITANTES POR FARMACIA SEGÚN LOS CRITERIOS DE ESTABLECIMIENTO..... | 23 |
| TABLA 8. COMPARATIVA DEL NÚMERO DE FARMACIAS Y HABITANTES POR FARMACIA SEGÚN SI LOS MEDICAMENTOS SIN RECETA SÓLO SE VENDEN EN LA FARMACIA | 24 |
| TABLA 9. PARTICIPACIÓN EN EL PAGO DE MEDICAMENTOS EN PAÍSES DE LA UE27 | 26 |
| TABLA 10. VENTAS TOTALES EN MILLONES DE EUROS. A PRECIOS PVP EXCEPTO LOS MARCADOS CON * (PRECIO DE VENTA DEL FABRICANTE) Y CON ** (PRECIO AL POR MAYOR) | 28 |
| TABLA 11. TIPO Y PESO DE LA DISTRIBUCIÓN MAYORISTA DE GAMA COMPLETA..... | 38 |
| TABLA 12. COMPOSICIÓN RELATIVA DEL PRECIO (PVP) DE LOS MEDICAMENTOS FINANCIADOS POR EL SNS Y PARTICIPACIÓN EN EL MISMO DE CADA AGENTE DEL SECTOR..... | 41 |
| TABLA 13. MÁRGENES DE LOS MEDICAMENTOS DE DISTRIBUCIÓN MAYORISTA Y FARMACIAS | 42 |
| TABLA 14. NÚMERO DE FARMACIAS Y RATIO DE HABITANTES/FARMACIA POR COMUNIDADES AUTÓNOMAS | 46 |

| | |
|---|----|
| TABLA 15. EVOLUCIÓN DEL NÚMERO DE COLEGIADOS Y TITULARES DE FARMACIA POR SEXO EN ESPAÑA EN EL PERÍODO (2005-2013) | 47 |
| TABLA 16. GASTO SANITARIO PÚBLICO CONSOLIDADO. MILLONES DE EUROS, PORCENTAJE SOBRE PIB Y EUROS POR HABITANTE. ESPAÑA, 2008-2012 | 51 |
| TABLA 17. GASTO SANITARIO PÚBLICO CONSOLIDADO SEGÚN CLASIFICACIÓN FUNCIONAL. MILLONES DE EUROS. ESPAÑA, 2008-2012 | 52 |
| TABLA 18. CUOTA SOBRE EL GASTO SANITARIO PÚBLICO CONSOLIDADO SEGÚN CLASIFICACIÓN FUNCIONAL. ESPAÑA, 2008-2012 | 53 |
| TABLA 19. ALTERNATIVAS EN LA REGULACIÓN ECONÓMICA DE LOS MEDICAMENTOS..... | 57 |
| TABLA 20. COMPARACIÓN DEL SEGMENTO DE GENÉRICOS DEL AÑO 2014 RESPECTO DEL 2013 EN UNIDADES Y VALORES POR COMUNIDADES AUTÓNOMAS | 63 |
| TABLA 21. TASA DE VARIACIÓN INTERANUAL (DICIEMBRE 2014 RESPECTO DICIEMBRE 2013)..... | 64 |
| TABLA 22. INDICADORES DEL GASTO FARMACÉUTICO EN EL PERIODO 2008-2014 A NIVEL NACIONAL .. | 75 |
| TABLA 23. NÚMERO DE RECETAS FACTURADAS DEL SISTEMA NACIONAL DE SALUD EN EL PERIODO 2008-2014.TASAS DE VARIACIÓN Y REPARTO SEGÚN LAS CCAA | 78 |
| TABLA 24. NÚMEROS ÍNDICE, BASE=2008, DEL NÚMERO DE RECETAS FACTURADAS DEL SISTEMA NACIONAL DE SALUD | 80 |
| TABLA 25. REGRESIÓN LINEAL ENTRE LAS CUOTAS DE PARTICIPACIÓN SOBRE EL TOTAL DE RECETAS FACTURADAS Y POBLACIÓN..... | 82 |
| TABLA 26. GASTO FARMACÉUTICO A TRAVÉS DE RECETA OFICIAL DEL SISTEMA NACIONAL DE SALUD EN EL PERIODO 2008-2014.TASAS DE VARIACIÓN Y REPARTO SEGÚN LAS CCAA | 83 |

| | |
|---|-----|
| TABLA 27. NÚMEROS ÍNDICE, BASE=2008, DEL GASTO FARMACÉUTICO A TRAVÉS DE RECETA OFICIAL DEL SISTEMA NACIONAL DE SALUD..... | 85 |
| TABLA 28. REGRESIÓN LINEAL ENTRE LAS CUOTAS DE PARTICIPACIÓN SOBRE EL TOTAL DEL GASTO EN RECETAS Y POBLACIÓN..... | 87 |
| TABLA 29. GASTO MEDIO POR RECETA FACTURADA DEL SISTEMA NACIONAL DE SALUD EN EL PERIODO 2008-2014.TASAS DE VARIACIÓN Y REPARTO SEGÚN LAS CCAA..... | 88 |
| TABLA 30. ESTRUCTURA DEL MERCADO DEL AUTOCUIDADO..... | 90 |
| TABLA 31. EVOLUCIÓN DEL MERCADO DE AUTOCUIDADO EN VALORES (PVP €) Y UNIDADES. 2010-2014 | 94 |
| TABLA 32. EVOLUCIÓN DEL MERCADO DE AUTOCUIDADOS, SEGÚN SEGMENTOS, EN VALORES (PVP €) Y UNIDADES. 2010-2014 | 95 |
| TABLA 33. RESULTADO ECONÓMICO-FISCAL DE LAS OFICINAS DE FARMACIA 2013 (BASE: DECLARACIONES IRPF / PATRIMONIO 2.012)..... | 101 |
| TABLA 34. CLASIFICACIÓN DE LOS SERVICIOS EN UNA OFICINA DE FARMACIA | 120 |
| TABLA 35. OTROS INDICADORES DE FIDELIZACIÓN | 133 |
| TABLA 36. CÁLCULO DEL DPP..... | 135 |
| TABLA 37. DATOS INICIALES SIMULACIÓN CLV | 136 |
| TABLA 38. COMPARACIÓN DEL CLV CON Y SIN CAMPAÑA | 137 |
| TABLA 39. BENEFICIOS DE LA FIDELIZACIÓN PARA CLIENTES/PACIENTES Y FARMACIAS | 140 |
| TABLA 40. VARIABLES DE SEGMENTACIÓN DE MERCADOS UTILIZADAS EN LOS MERCADOS DE CONSUMO | 150 |

| | |
|--|-----|
| TABLA 41. ESTRATEGIAS BÁSICAS DE UNA EMPRESA PARA SELECCIONAR EL MERCADO META. | |
| P=PRODUCTO/SERVICIO, M=MERCADO..... | 152 |
| TABLA 42. VALORES DE K PARA DIFERENTES TAMAÑOS DE VOLUMEN DE DATOS | 164 |
| TABLA 43. PRINCIPALES FUNCIONES DISTANCIA..... | 169 |
| TABLA 44. CONCEPTOS BÁSICOS SOBRE CLUSTERING | 179 |
| TABLA 45. MÉTODOS Y ALGORITMOS DE SEGMENTACIÓN: CLASIFICACIÓN AMPLIADA | 183 |
| TABLA 46. IMPACTO DE LA EFICIENCIA DE UN ALGORITMO..... | 185 |
| TABLA 47. EJEMPLO DE CRUCE CON OPERADOR BINOMIAL | 225 |
| TABLA 48. CARACTERÍSTICAS DE LAS VARIABLES DEL CONJUNTO DE DATOS DE LA OFICINA DE FARMACIA | 259 |
| TABLA 49. RESULTADOS OBTENIDOS PARA EL CONJUNTO S1 | 273 |
| TABLA 50. RESULTADOS OBTENIDOS PARA EL CONJUNTO S2 | 274 |
| TABLA 51. RESULTADOS OBTENIDOS PARA EL CONJUNTO S2 | 275 |
| TABLA 52. RESULTADOS COMPARATIVOS ENTRE ACDE SIN REORDENAMIENTO Y ACDE CON REORDENAMIENTO Y CON LA CODIFICACIÓN PROPUESTA EN PROBLEMAS DE TAMAÑO MEDIO..... | 277 |
| TABLA 53. RESULTADOS COMPARATIVOS ENTRE ACDE SIN REORDENAMIENTO Y ACDE CON REORDENAMIENTO Y CON LA CODIFICACIÓN PROPUESTA EN PROBLEMAS DE GRAN TAMAÑO | 281 |
| TABLA 54. RESULTADOS DEL CONTRASTE KRUSKAL-WALLIS PARA LOS CUATRO TAMAÑOS DEL MUESTREO BASADO EN HIPERCUBOS LATINOS PARA EL PROBLEMA IRIS | 291 |
| TABLA 55. RESULTADOS DEL CONTRASTE KRUSKAL-WALLIS PARA LOS CUATRO TAMAÑOS DEL MUESTREO BASADO EN HIPERCUBOS LATINOS PARA EL PROBLEMA GLASS | 294 |

| | |
|--|-----|
| TABLA 56. RESULTADOS DEL CONTRASTE KRUSKAL-WALLIS PARA LOS CUATRO TAMAÑOS DEL MUESTREO BASADO EN HIPERCUBOS LATINOS PARA EL PROBLEMA S1..... | 297 |
| TABLA 57. RESULTADOS DEL CONTRASTE KRUSKAL-WALLIS PARA LOS CUATRO TAMAÑOS DEL MUESTREO BASADO EN HIPERCUBOS LATINOS PARA EL PROBLEMA YEAST | 300 |
| TABLA 58. PARÁMETROS DE EJECUCIÓN PARA LOS ALGORITMOS ACDE Y DECCS | 301 |
| TABLA 59. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN IRIS..... | 303 |
| TABLA 60. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN GLASS | 304 |
| TABLA 61. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN WINE | 305 |
| TABLA 62. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN BREAST | 306 |
| TABLA 63. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN THYROID | 307 |
| TABLA 64. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN LED..... | 308 |
| TABLA 65. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN YEAST..... | 309 |
| TABLA 66. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN WINE QUALITY..... | 310 |
| TABLA 67. RESULTADOS COMPARATIVOS ENTRE LOS CINCO ALGORITMOS EN SONAR | 311 |
| TABLA 68. COMPARACIÓN ENTRE ACDE Y DECCS | 312 |
| TABLA 69. COMPARACIÓN ENTRE ACDE Y DECCS CON UMBRAL DE PARADA | 313 |
| TABLA 70. RESULTADOS PARA ACDE EN EL CONJUNTO PEN-BASED RECOGNITION OF HANDWRITTEN DIGITS | 314 |
| TABLA 71. RESULTADOS PARA DECCS EN EL CONJUNTO PEN-BASED RECOGNITION OF HANDWRITTEN DIGITS | 314 |
| TABLA 72. SEGMENTACIÓN OBTENIDA EN LA APLICACIÓN DE DECCS EN UNA OFICINA DE FARMACIA.. | 316 |
| TABLA 73. PREGUNTA 1: SU TIPO DE FARMACIA ES..... | 321 |

| | |
|--|-----|
| TABLA 74. PREGUNTAS 1 Y 2: TIPO DE FARMACIA SEGÚN SI ES LA ÚNICA O NO EN EL MUNICIPIO. | |
| FRECUENCIAS ABSOLUTAS..... | 321 |
| TABLA 75. PREGUNTAS 1 Y 2: TIPO DE FARMACIA SEGÚN SI ES LA ÚNICA O NO EN EL MUNICIPIO. | |
| PORCENTAJES SOBRE EL TOTAL..... | 321 |
| TABLA 76. PREGUNTAS 1 Y 3: TIPO DE FARMACIA SEGÚN SI LA DISTANCIA AL CENTRO DE SALUD MÁS | |
| PRÓXIMO ES INFERIOR A 250 METROS. FRECUENCIAS ABSOLUTAS..... | |
| | 322 |
| TABLA 77. PREGUNTAS 1 Y 3: TIPO DE FARMACIA SEGÚN SI LA DISTANCIA AL CENTRO DE SALUD MÁS | |
| PRÓXIMO ES INFERIOR A 250 METROS. PORCENTAJES SOBRE EL TOTAL..... | |
| | 323 |
| TABLA 78. PREGUNTAS 4 Y 7: DISTINCIÓN CLIENTES Y PACIENTES SEGÚN SI LA OFICINA DE FARMACIA POSEE | |
| O NO TARJETA DE FIDELIZACIÓN. FRECUENCIAS ABSOLUTAS..... | |
| | 324 |
| TABLA 79. PREGUNTAS 4 Y 7: DISTINCIÓN CLIENTES Y PACIENTES SEGÚN SI LA OFICINA DE FARMACIA POSEE | |
| O NO TARJETA DE FIDELIZACIÓN. PORCENTAJES SOBRE EL TOTAL..... | |
| | 324 |
| TABLA 80. PREGUNTAS 2,4 Y 7: DISTINCIÓN CLIENTES Y PACIENTES SEGÚN SI LA OFICINA DE FARMACIA | |
| POSEE O NO TARJETA DE FIDELIZACIÓN Y SI LA OFICINA DE FARMACIA ES ÚNICA O NO EN SU LOCALIDAD. | |
| FRECUENCIAS ABSOLUTAS..... | 325 |
| TABLA 81. PREGUNTAS 2,4 Y 7: DISTINCIÓN CLIENTES Y PACIENTES SEGÚN SI LA OFICINA DE FARMACIA | |
| POSEE O NO TARJETA DE FIDELIZACIÓN Y SI LA OFICINA DE FARMACIA ES ÚNICA O NO EN SU LOCALIDAD. | |
| PORCENTAJES SOBRE EL TOTAL..... | 325 |
| TABLA 82. PREGUNTA 5: ¿QUÉ MEDIOS UTILIZA PARA COMUNICARSE CON SUS CLIENTES/PACIENTES? | |
| | 327 |
| TABLA 83. PREGUNTA 6: ¿CUÁL ES EL MEDIO QUE MÁS UTILIZA? | |
| | 328 |
| TABLA 84. PREGUNTA 7: ¿QUÉ INFORMACIÓN REGISTRA DE SUS CLIENTES/PACIENTES? | |
| | 329 |

1 INTRODUCCIÓN Y ENTENDIMIENTO DEL PROBLEMA

Si consideramos que innovar no es inventar algo que nunca ha existido sino desprenderse de las ideas preconcebidas, esta investigación pretende vaciar la mochila de clichés y estereotipos más o menos fundados y abordar la utilización de herramientas por parte de las oficinas de farmacia considerando su entorno, su trayectoria, su situación actual y sus posibles escenarios de futuro. Por supuesto, este camino no es único. Existen muchas alternativas y este trabajo pretende abordar alguna de ellas.

En un sector tan complejo como lo es el farmacéutico, sería una temeridad no pararse a analizarlo con la suficiente profundidad y delicadez como para que las técnicas y herramientas aquí propuestas, no tuvieran visos de ser “implantables”. Particularmente, y a través de los años de experiencia acumulados en mi carrera profesional, he acuñado el término “implantabilidad en tiempo de diseño” queriendo hacer referencia a que todo aquello que propongamos, independiente de lo sesudo del estudio, debe ser “implantable”, o lo que es lo mismo, adoptado por el usuario sin que medie un cambio radical en los procesos de negocio de los que participa. En definitiva, ningún sistema o herramienta debería provocar un cambio radical en la forma de hacer las cosas.

La oficina de farmacia, y todo su entorno, está viviendo uno de sus momentos más convulsos de las últimas décadas y no parece oportuno quedarse inmóvil frente a una más que posible evolución del sector. La farmacia ha sido un negocio muy estable desde sus inicios y el colectivo farmacéutico español, unido como ninguno, se está viendo sometido a una sacudida provocada por las amenazas sobre su modelo y sus oportunas consecuencias negativas. El mercado regulado del medicamento ha visto como el volumen de facturación ha experimentado una importante contracción provocada por las medidas de contención del gasto farmacéutico tomadas por la Administración como consecuencia de la situación de crisis por la que está pasando España.

Analizando los puntos fuertes de la oficina de farmacia, que son muchos, y enfrentándolos con la estrecha relación que mantienen con sus usuarios (clientes y/o pacientes) se genera una oportunidad, que se torna alternativa a la tensa situación descrita, para abordar el mercado del autocuidado de la salud y la venta libre implantando para ello técnicas y herramientas que tengan como ingrediente estrella a

la caracterización o segmentación de sus clientes. Así mismo, el uso de los sistemas de información para la gestión de las relaciones con sus clientes debe potenciarse como medida de rentabilidad y diferenciación respecto a los establecimientos no sanitarios que compiten de forma nativa en el mercado del autocuidado de la salud.

Por otro lado, y como dijo Einstein, “si siempre haces lo mismo siempre obtendrás los mismos resultados” y en el mercado global en el que hoy nos relacionamos, donde el cambio es lo habitual, y en estos momentos donde en el sector de la farmacia se están presentando tantas turbulencias, debemos plantearnos nuevos caminos, sin olvidar y mejorar nuestro desempeño en los actuales, que nos permitan reducir la incertidumbre ante la aparición de nuevos escenarios de futuro.

Para abordar esta investigación, el documento se ha estructurado en seis secciones principales que van desde lo general o macro, materializado en un análisis del sector, evolución y sus agentes, a lo específico, concreto o micro, plasmado en técnicas y herramientas concretas para la mitigación del riesgo de una posible situación desfavorable en el negocio de las oficinas de farmacia.

En la primera sección, en la que nos encontramos, se especifican los objetivos de esta investigación. Bajo el título “El entorno de la oficina de farmacia” se describe el entorno de la oficina de farmacia y se realiza un análisis pormenorizado de su situación actual. Además en este capítulo se analizan sus agentes y la relación entre ellos.

A continuación, “Herramientas para el desarrollo de la oficina de farmacia” que presenta un conjunto de conceptos, técnicas y herramientas que ayudan al desarrollo de las oficinas de farmacia en un mercado en el que no están acostumbradas a competir mientras que los competidores a los que se enfrentan, parafarmacias y grandes almacenes, son expertos en el uso de dichas herramientas.

Por último en esta sección, y una vez desarrollados los conceptos previos, se aborda “El problema estadístico de la segmentación”. En el que se estudia la segmentación desde el punto de vista estadístico. En punto se comparan los conceptos de segmentación y clasificación, cuyas diferencias son notables a la hora de plantear un enfoque u otro.

También se analizan los pros y contras en el uso de las bases de datos y los aspectos clave a tener en cuenta antes de aplicar una técnica de *clustering*: determinación del número de clústeres, indicadores de validez, tipos de algoritmos y variantes. Para finalizar, se estudian con detalle los algoritmos metaheurísticos de segmentación.

En la segunda sección se describe la metodología seguida en la investigación. En primer lugar, se especifica el algoritmo de segmentación bioinspirado propuesto, mostrándose las variaciones respecto al original y que conforman una de las aportaciones de la investigación. Asimismo, se da solución a los problemas de semillado y aplicación a las grandes bases de datos. A continuación, se describen los conjuntos sobre los que se ejecutarán los algoritmos, el software y hardware utilizado y la encuesta realizada a los farmacéuticos de Sevilla.

En la tercera sección, se ofrecen los resultados y su discusión. Se e los resultados obtenidos por la aplicación tanto del algoritmo original como el propuesto en los conjuntos de datos. Incluyendo los resultados obtenidos sobre un conjunto de datos propio de una oficina de farmacia. En la última parte de esta sección, se analizan los resultados obtenidos de la encuesta realizada a los farmacéuticos.

En la cuarta sección se refleja el resumen de la investigación y sus conclusiones. Así como posibles líneas de investigación a seguir para profundizar en aspectos de interés en el problema de segmentación y su aplicación en las oficinas de farmacia.

En la quinta sección se muestran todas las fuentes bibliográficas consultadas en esta investigación.

En la sexta y última sección, se establecen los anexos con información complementaria de utilidad en la investigación.

1.1 Objetivos

A partir de la situación descrita y de los argumentos planteados, el principal objetivo de esta investigación es encontrar un modelo de caracterización de los usuarios (clientes y pacientes) de una oficina de farmacia en base a variables discriminantes y a un algoritmo de segmentación adecuado y contrastado. De esta forma, las oficinas de farmacia que lo implanten podrán tomarlo como base para la puesta en marcha de planes de fidelización ajustados a las necesidades detectadas en sus segmentos de clientes.

Como objetivos de segundo nivel que se persiguen en esta investigación nos proponemos los siguientes:

- Caracterización del sector de la farmacia en España.
- Configurar el mapa de los agentes del sector farmacéutico en España, su relación y su situación en el momento actual.
- Análisis de la evolución del mercado farmacéutico español
- Diseñar e implementar un algoritmo de segmentación flexible, eficiente (en calidad de soluciones y tiempos) y robusto, capaz de resolver problemas de grandes dimensiones.
- Aportar una definición para pacientes y clientes de una oficina de farmacia
- Estudiar las características de grupos homogéneos de clientes
- Encontrar un patrón de comportamiento de consumo para cada uno de los grupos definidos de clientes.
- Identificar futuras líneas de investigación en la definición de herramientas basadas en procedimientos estadísticos para las oficinas de farmacia.

Se propone para la segmentación de clientes el uso de técnicas estadísticas de *clustering* basadas en algoritmos metaheurísticos, previa comparación con otro tipo de algoritmos, esperando que el resultado sea satisfactorio en la segmentación de los usuarios.

Otro propósito destacado de esta investigación es estudiar las posibles herramientas en lo que respecta a la relación de las oficinas de farmacia con sus clientes.

Partiendo de los datos obtenidos en una oficina de farmacia se aplicará el algoritmo propuesto para la obtención de segmentos de clientes y así caracterizarlos de forma que el farmacéutico pueda utilizar dicha caracterización para rentabilizar su negocio.

Esta investigación permitirá sentar las bases del uso de técnicas estadísticas en las actividades comerciales y de marketing en las oficinas de farmacia.

1.2 El entorno de la oficina de farmacia

En línea con los objetivos fijados en esta tesis, conocer en qué entorno se desenvuelve la oficina de farmacia merece dedicar un apartado específico. Por ello, abundaremos en el conocimiento de los modelos farmacéuticos más representativos a nivel internacional, el modelo farmacéutico español, su sistema sanitario, el mercado farmacéutico, sus agentes, sus parámetros principales y aquellos otros aspectos que configuran la actividad de la oficina de farmacia. Y ésta, tiene un papel protagonista reservado dentro de la cadena de valor del sector sanitario (1) como se recoge en la Figura 1.

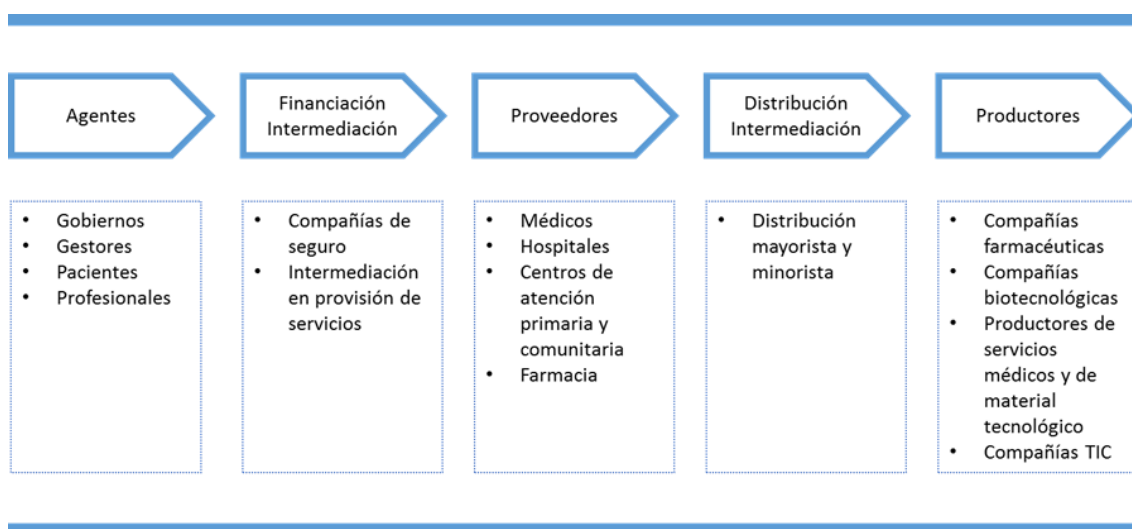


Figura 1. Cadena de valor del sector sanitario

1.2.1 Los modelos farmacéuticos

Según se observa en los estudios sectoriales realizados, las variables que definen un modelo farmacéutico son las tres que siguen por orden de relevancia (2):

1. Binomio propiedad/titularidad
2. Planificación geográfica o demográfica
3. La venta de medicamentos sin receta

Desde un punto de vista internacional, y según los elementos y las características que los determinan, existen dos tipos de modelos de farmacia bien diferenciados como se indica en la Figura 2.

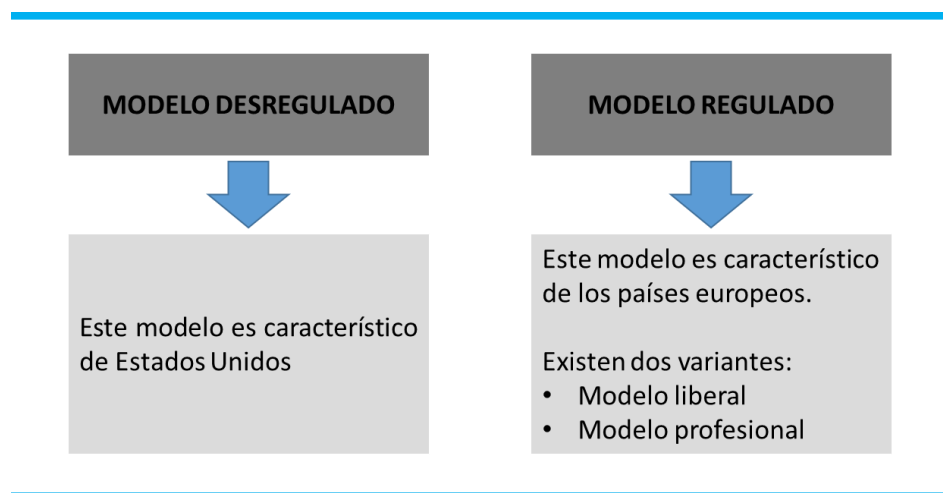


Figura 2. Modelos de farmacia

Para comprender las diferencias entre ambos es conveniente tener en cuenta el sistema sanitario en el que tienen lugar. Por ello, entendemos necesario describir previamente los diferentes modelos de sistemas sanitarios.

Los modelos sanitarios

Los modelos sanitarios (3) se basan en la combinación y equilibrio entre tres grandes funciones que los definen:

- Regulación
- Financiación
- Provisión

Si analizamos los distintos tipos de financiación podemos identificar cuatro grandes modelos de sistemas sanitarios (4):

- Sistema del estado de bienestar (modelo Beverige y Shemasko)
 - Servicios financiados por impuestos
 - Control estatal
- Seguridad Social (modelo Bismarck)
 - Cotizaciones
 - Participación obligatoria de los trabajadores
- Aseguramiento voluntario (privado)
 - La participación depende de la decisión individual
- Pago directo de los servicios

En la tabla que se recoge a continuación se puede apreciar un esquema de los distintos tipos de sistemas sanitarios y los países que disponen de ellos, según las funciones anteriormente comentadas:

| Funciones | SNS | Seguridad Social | Privados |
|---------------|--|--------------------------------------|----------|
| Financiación | Pública (impuestos) | Pública (cotizaciones salariales) | Privada |
| Aseguramiento | Pública (universal) | Pública (múltiple) | Privada |
| Provisión | Pública | Privada | Privada |
| Países | Países Nórdicos Reino Unido España | Centroeuropa Francia Suiza | EEUU |

Tabla 1. Tipos de sistemas sanitarios según funciones y países

Si observamos con detenimiento los sistemas sanitarios de los países ocurre que en la mayor parte de ellos, los sistemas de financiación y aseguramiento no son puros en sí mismos aunque de forma habitual se toma la característica fundamental como definitoria (5). Se distinguen dos grupos de países: uno, fundamentalmente europeo, caracterizado por tener Sistemas Nacionales de Salud y de Seguridad Social, y un segundo, que se decanta por un modelo de aseguramiento privado; grupo que está encabezado por EEUU (4).

Por todo ello podemos observar como los sistemas sanitarios europeos son opuestos al sistema sanitario de EEUU y esto en cierta medida explica la diferencia existente entre los modelos farmacéuticos de Europa y EEUU que describiremos más adelante.

En un estudio comparativo de los sistemas sanitarios (6) evaluando las distintas dimensiones de los mismos como el nivel de salud, satisfacción, gasto, etc., se determina que los Sistemas Nacionales de Salud son los que aportan mayor eficiencia global.

No obstante, y según el informe realizado por PriceWaterhouseCoopers (7), se está produciendo una convergencia entre los sistemas de seguro de Estados Unidos y Europa.

En Europa, donde la sanidad es un derecho de todos los ciudadanos, el seguro privado cubre alrededor del 10% de los costes sanitarios y crece entre un 5% y un 7% cada año. En Estados Unidos, donde la sanidad es un beneficio de los trabajadores por el que pagan las empresas, el gasto público ha aumentado en 10 años del 40% a casi el 50% y

continúa el debate sobre la extensión de los beneficios públicos a todos los ciudadanos. El estudio sugiere que finalmente se impondrán sistemas de seguro básico y obligatorio, y seguros adicionales u opcionales, sufragados por el consumidor o por las empresas.

El modelo sanitario español, considerado como el séptimo mejor a nivel mundial por la Organización Mundial de la Salud (OMS) en el año 2000, se basa en tres principios básicos:

- La cobertura universal.
- Financiación pública.
- Gestión descentralizada.

Sus características principales son las siguientes (3):

1. Financiado por impuestos.
2. Cobertura universal.
3. Amplia cartera pública de cobertura sanitaria.
4. Copago de la prestación farmacéutica.
5. Provisión mayoritaria de proveedores públicos.
6. Gestión descentralizada.
7. Organización por niveles asistenciales.

Dado el interés que tiene para esta investigación, entendemos necesario detallar el copago (porcentaje de aportación del usuario en el pago de medicamentos prescritos) de la prestación farmacéutica que presenta el modelo español que se expone en la Tabla 2. Máxime cuando se vio modificado en el año 2012, con el Real Decreto-ley 16/2012, de 20 de abril, de medidas urgentes para garantizar la sostenibilidad del Sistema Nacional de Salud y mejorar la calidad y seguridad de sus prestaciones (8), produciendo un impacto considerable en la facturación de los medicamentos de prescripción en las oficinas de farmacia (9).

| Pago de medicamentos. Aportación del usuario (asegurado y beneficiarios) | | | |
|--|---------------------------------------|-------------------|---|
| Situación | Antes de aplicar el RD 16/2012 | En general | A partir de 01/07/2012 Límite máximo de aportación para el tratamiento de larga duración (€/mes) a enero de 2015 |
| Pensionistas con renta inferior a 18.000€/año. | 0% | 10% | 8,23 € |
| Pensionistas con renta entre 18.000 y 100.000€/año | 0% | 10% | 18,52 € |
| Pensionistas con renta igual o superior a 100.000€/año | 0% | 60 % | 61,75 € |
| Personas sin recursos que perciben renta de integración | 40% | 0% | 0 € |
| Personas perceptoras de pensiones no contributivas | 0% | 0% | 0% |
| Afectados del síndrome tóxico y personas con discapacidad en los supuestos contemplados en su normativa específica | 0% | 0% | 0 € |
| Parados que han perdido el subsidio de desempleo | 40% | 0% | 0 € |
| Personas con tratamientos por accidente de trabajo o enfermedad profesional | 0% | 0% | 0 € |
| Usuarios con tratamientos crónicos para los medicamentos pertenecientes a los grupos ATC y productos sanitarios de aportación reducida | 10 %, con límite de 2,64€ | 10% | Límite por receta: 4,26€ |
| Usuarios con renta inferior a 18.000€/año | 40% | 40% | Sin límite |
| Usuarios con renta entre 18.000€/año y 100.000€/año | 40% | 50% | Sin límite |
| Usuarios con renta igual o superior a 100.000€/año | 40% | 60% | Sin límite |
| Mutualistas y clases pasivas de MUFACE, ISFAS y MUGEJU | 30% | 30% | Sin límite |

Tabla 2. Copago antes y después del RD 16/2012

Como veremos más adelante, con la implantación del citado Decreto (9), se produjo un descenso en el número de recetas del 12,18%, y de un 12,83% en el caso del gasto farmacéutico. Este, junto con otros que se verán más adelante, es uno de los hechos que justifican los objetivos de esta investigación.

1.2.1.1 Los modelos farmacéuticos de Estados Unidos y Japón

Con el objeto de disponer de otros modelos diferentes al caso europeo, y disponer de una cierta perspectiva, a continuación se describen los modelos farmacéuticos de Estados Unidos y Japón.

El modelo farmacéutico de Estados Unidos

El modelo farmacéutico estadounidense, que se desarrolla en un sistema liberal de economía de mercado, presenta grandes diferencias con respecto al modelo europeo en sus dos variantes (liberal y profesional).

Como hemos visto anteriormente, es obligado observar los sistemas sanitarios donde se desarrollan los modelos estadounidense y europeo para poder comprender las diferencias entre el modelo de farmacia de cada uno. Estas diferencias están principalmente relacionadas con aspectos legislativos que, a su vez, condicionan la política sanitaria que sigue Estados Unidos o los países miembros de la Unión Europea.

Estados Unidos no posee un sistema nacional de sanidad pública centralizado, como el predominante que existe en Europa. La estructura estadounidense de salud está formada por dos grandes sistemas: Medicare, que desde 1966 está destinado a un segmento de población mayor de 65 años o que se encuentra en situación de desempleo por un período superior a 2 años, y que desde 1973 incluye también a pacientes en estado final de enfermedad renal; y Medicaid, que es un seguro médico destinado a personas de rentas muy bajas. La población no incluida en alguno de estos dos sistemas estatales debe costearse su atención sanitaria. Y dado el elevado coste de la sanidad privada, lo habitual es que los estadounidenses contraten un seguro de cobertura a través de una aseguradora privada.

El funcionamiento de estas aseguradoras es similar al caso español. Es un médico de atención primaria el que tras una primera exploración remite el paciente al especialista o bien es el paciente el que puede dirigirse directamente al especialista que desee. En el sistema estadounidense, en ambos casos, una vez se ha realizado la visita, el paciente debe pagar una cuota de asistencia.

Con respecto a la adquisición de medicamentos, a diferencia de lo que ocurre en Europa, existen diferentes lugares en los que los ciudadanos estadounidenses pueden adquirirlos (10):

- Drugstores.
- Farmacias independientes.
- Supermercados.
- Correo (Internet o correo postal).

El drugstore, que ocupa el primer lugar en la cuota de mercado, nació en 1929 y supuso una gran revolución para la época con su diseño innovador. El drugstore suponía un nuevo concepto de espacio utilizado para la venta, con precios justos y una gran exposición de productos complementarios y alternativos al medicamento. Son establecimientos con una exposición masiva de productos para que cada comprador elija los fármacos que más les interesen o llamen su atención. Los drugstores, repartidos por todo el país, están agrupados en cadenas y asociados en la Asociación Nacional de Cadenas de Drugstores (NACDS).

Las farmacias independientes, que ocupan el segundo lugar en cuota de mercado, son aquellas que no son propiedad de ningún grupo de drugstores y suelen pertenecer a un propietario. Este no tiene la obligatoriedad de ser farmacéutico y puede poseer más de una farmacia. Lo que si es necesario es que en las farmacias independientes el responsable sea un Licenciado en Farmacia. Por sus características, es el establecimiento que más se asemeja a la oficina de farmacia europea.

La extrema competencia entre las cadenas de drugstores y las farmacias independientes ha provocado que en estas últimas puedan encontrarse objetos que poco tienen que ver

con la salud (como cosmética de lujo, juguetes, complementos de lujo para animales de compañía y aparatos de electrónica de consumo).

La venta de medicamentos por correo en EEUU tiene su origen en el año 1995 y fue muy bien acogida por los jubilados estadounidenses quienes a través del sistema de correo tradicional y enviando las prescripciones a una serie de establecimientos determinados se evitaban pagar el copago de las medicinas. Los jubilados deben formar parte de un programa sanitario y las recetas deben enviarse a un centro destinado a la compra por correo de medicamentos. Sólo de esta forma pueden recibir los medicamentos en su casa sin necesidad de desplazarse a la farmacia de forma presencial.

Los supermercados de EEUU tienen la posibilidad de disponer en su local de una sección de farmacia. En estos establecimientos la compra de medicamentos OTC (medicamento de venta libre o medicamento sin prescripción llamado así por sus siglas en inglés *Over The Counter*) es relevante en volumen y se da una situación de choque, al menos respecto a la farmacia en España, al encontrar medicamentos para la automedicación junto a pasteles, salsa de tomate o patatas chips (10).

En los supermercados y grandes superficies, la automedicación adquiere un papel protagonista, que se ve acentuado por los distintos embalajes, cada cual más atractivo y sugerente a la vista.

En lo que respecta al sistema de distribución de medicamentos, el funcionamiento de la farmacia estadounidense es muy parecido al europeo. Existen unos distribuidores mayoristas de medicamentos que tramitan varias veces al día los pedidos solicitados por las farmacias.

Sin embargo, en lo referente a los consumos de productos OTC, las diferencias son importantes. En Estados Unidos, al existir la automedicación, el usuario dedica mucho tiempo a analizar los productos y presentaciones antes de comprar la medicación. La presencia del farmacéutico está destinada a trámites más técnicos y muchas veces la consulta se solicita con cita previa.

El modelo farmacéutico japonés

El modelo japonés es similar al de Estados Unidos. En este caso, los establecimientos a los que los japoneses se dirigen para adquirir medicamentos son los siguientes (11):

| Tipo de establecimiento | % de establecimientos sobre el total |
|---|--------------------------------------|
| Farmacia con farmacéutico | 51,5% |
| Droguería con farmacéutico | 12,5% |
| Droguería sin farmacéutico | 14,9% |
| Tokurei-hanbai-gyo (licencia que se concede cuando en un lugar no hay ni farmacia ni droguería. Limitada a una serie de fármacos) | 9,7% |
| Haichi-hanbai-gyo (distribuidos por el propietario. Limitada a una serie de fármacos) | 11,4% |

Tabla 3. Establecimientos de venta de medicamentos en Japón

Los fármacos sin prescripción médica debían ser vendidos por un farmacéutico hasta 2004. Tras la modificación de la ley en el año 2004, 371 productos de este tipo fueron aprobados para su distribución en tiendas de 24 horas o en supermercados. Entre estos fármacos se incluyen remedios gastrointestinales, algunos laxantes, tónicos vitamínicos o productos con calcio.

En el caso europeo que tratamos a continuación, y en base al análisis de los estudios e informes consultados en esta investigación sobre los modelos farmacéuticos de la UE, puede observarse un panorama general en el que se observa la existencia de, al menos tres modelos, según las variables binomio propiedad/titularidad, planificación geográfica o demográfica y venta de medicamentos sin receta.

1.2.1.2 Los modelos farmacéuticos europeos

Las distintas realidades sanitarias y sociales de los países europeos y sus regímenes de protección social, determinan las diferencias en la ordenación farmacéutica de cada modelo.

Si bien es cierto que los estudios sectoriales resumen los modelos de farmacia europeos en dos grandes grupos (profesional y liberal), creemos necesario afinar más esta clasificación para poder mejorar nuestro punto de vista en el análisis propuesto.

Como conclusiones de un comité de expertos (12), los modelos considerados según el criterio de planificación de las oficinas de farmacia son los siguientes:

- Mediterráneo: este modelo regula por ley los criterios de apertura para asegurar una cobertura homogénea en su territorio y de fácil acceso de la población a la farmacia. La decisión sobre el número y localización de las oficinas de farmacia se resuelve, para este modelo, conforme a criterios que tienen en cuenta módulos de población y de distancia entre oficinas de farmacia.
- Continental: este modelo presenta dos alternativas opuestas. Por un lado, están Alemania y Luxemburgo, reticentes a imponer condicionantes a la apertura de farmacias, aunque el rendimiento económico y la demanda son los criterios respectivos para la apertura de las mismas. Por otro lado, la otra alternativa del modelo viene representada por Francia, Bélgica y Austria, con módulos de población como requisito (criterio poblacional).
- Nórdico: este modelo no impone restricciones legales, en la práctica es el Estado o las propias asociaciones farmacéuticas (caso de Suecia) las que marcan limitaciones en el número y localización. La escasa densidad poblacional de sus territorios obliga a las autoridades a planificar la apertura de farmacias de una forma ordenada que beneficie al ciudadano.
- Anglosajón: representado por el Reino Unido y Holanda, de corte más liberal, ofrece más flexibilidad subrayando la naturaleza privada del ejercicio farmacéutico, aunque en la práctica este modelo no se traduce en una mayor cobertura farmacéutica, sino justo lo contrario.

Aunque hasta este año de 2015 ha habido turbulencias en la Unión Europea sobre los modelos expuestos, esta clasificación sigue vigente en la actualidad.

En la Tabla 4 podemos observar una comparación de los diferentes modelos de farmacia en la UE27, realizada por el Consejo General de Colegios Oficiales de Farmacéuticos (2), en base a si existe o no propiedad farmacéutica, si existen o no criterios de establecimiento y si se venden medicamentos fuera de las farmacias (13).

| País | Propiedad farmacéutica | Criterios de establecimiento demográficos y/o geográficos | Medicamentos sin receta sólo en farmacia |
|-----------------|------------------------|---|--|
| España | Sí | Sí | Sí |
| Finlandia | Sí | Sí | Sí |
| Francia | Sí | Sí | Sí |
| Grecia | Sí | Sí | Sí |
| Letonia | Sí | Sí | Sí |
| Luxemburgo | Sí | Sí | Sí |
| Austria | Sí | Sí | No |
| Dinamarca | Sí | Sí | No |
| Eslovenia | Sí | Sí | No |
| Italia | Sí | Sí | No |
| Alemania | Sí | No | Sí |
| Bulgaria | Sí | No | Sí |
| Chipre | Sí | No | Sí |
| Bélgica | No | Sí | Sí |
| Estonia | No | Sí | Sí |
| Malta | No | Sí | Sí |
| Portugal | No | Sí | No |
| Reino Unido | No | Sí (*) | No |
| Eslovaquia | No | No | Sí |
| Lituania | No | No | Sí |
| Rumania | No | No | Sí |
| Suecia | No | No (**) | Sí |
| Holanda | No | No | No |
| Hungría | No | No | No |
| Irlanda | No | No | No |
| Polonia | No | No | No |
| República Checa | No | No | No |

Tabla 4. Comparación de los diferentes modelos farmacéuticos en la UE27

* Para conseguir contrato con el Sistema Sanitario

** En el año 2009, el parlamento sueco procedió a la liberalización de la propiedad de las farmacias. Hasta ese momento la propiedad era estatal y las farmacias eran gestionadas por la Corporación Nacional Sueca de Farmacia (Apoteket AB). Esta medida afectó a 466 de sus farmacias quedando el resto en manos del Estado.

De esta comparación, se observa con claridad la existencia de tres grupos. El primero formado por España, Finlandia, Francia, Grecia, Letonia y Luxemburgo en el que las tres variables analizadas arrojan valores afirmativos. Al otro extremo surge un grupo con un modelo farmacéutico netamente liberal formado por Holanda, Hungría, Irlanda, Polonia y República Checa. Y finalmente, un grupo formado por los 16 países restantes con un modelo mixto.

Para mejorar el análisis de los modelos existentes en la UE27 combinaremos la información de la Tabla 4 con la información de la Tabla 5 en la que podemos observar una comparación del número de oficinas de farmacia y el ratio habitantes por farmacia en los países de la UE27.

| País | Nº de farmacias | Habitantes por Farmacia |
|---------------------|-----------------|-------------------------|
| España | 21.459 | 2.170 |
| Finlandia | 818 | 6.570 |
| Francia | 22.655 | 2.930 |
| Grecia | 11.000 | 1.028 |
| Letonia | 825 | 2.702 |
| Luxemburgo | 91 | 5.624 |
| Austria | 1.303 | 6.449 |
| Dinamarca | 314 | 17.708 |
| Eslovenia | 321 | 6.386 |
| Italia | 18.102 | 3.348 |
| Alemania | 20.921 | 4.100 |
| Bulgaria | 3.652 | 2.054 |
| Chipre | 477 | 1.686 |
| Bélgica | 4.998 | 2.190 |
| Estonia | 469 | 2.856 |
| Malta | 210 | 1.988 |
| Portugal | 2.885 | 3.686 |
| Reino Unido | 14.186 | 3.730 |
| Eslovaquia | 1.935 | 2.808 |
| Lituania | n.d | n.d |
| Rumania | 7.932 | 2.699 |
| Suecia | 1.303 | 7.225 |
| Holanda | 1.981 | 8.407 |
| Hungría | 2.334 | 4.277 |
| Irlanda | 1.701 | 2.634 |
| Polonia | 11.297 | 3.381 |
| República Checa | 2.450 | 4.298 |
| Total UE (*) | 155.619 | 3.169 |

Tabla 5. Comparativa del número de farmacias y población en la UE27

* Datos no disponibles de Lituania

Fuente: Número de farmacias: Agrupación Farmacéutica de la Unión Europea (PGEU)

Población: Eurostat Fecha: diciembre 2013

De la integración de ambas fuentes, podemos disponer de información de valor añadido para describir mejor las características de los modelos de la UE27 desde el punto de vista unidimensional.

En la Tabla 6 podemos observar que en el caso de los países donde la propiedad de la farmacia corresponde a un farmacéutico, el número de oficinas prácticamente dobla al de los países donde dicha restricción no se da. Además, presentan una ratio de habitantes por farmacia inferior en 635 habitantes. Esta característica está repartida en la misma proporción en la UE27. No obstante, estos países aglutinan la mayoría de las farmacias de la UE27 con un 65,5%.

| | Propiedad farmacéutica | | Número de farmacias | | Habitantes por farmacia |
|-----------|------------------------|------------|---------------------|------------|-------------------------|
| | Países | % sobre UE | Número | % sobre UE | |
| Sí | 13 | 48,15% | 101.938 | 65,50% | 2.950 |
| No | 14 | 51,85% | 53.681 | 34,50% | 3.585 |

Tabla 6. Comparativa del número de farmacias y habitantes por farmacia según la propiedad farmacéutica

En la Tabla 7 observamos que en el caso de los países donde existen criterios de establecimiento demográfico y/o geográfico, el número de oficinas es superior en más de un 75% al de los países donde dicha restricción no se da. Además presentan una ratio de habitantes por farmacia inferior en 848 habitantes. En este caso, la proporción de países que presentan esta característica en su modelo farmacéutico es superior a la proporción de los países que no la tienen. Asimismo, estos países aglutinan la mayoría de las farmacias de la UE27 con un 64,03%.

| | Criterios de establecimiento demográficos y/o geográficos | | Número de farmacias | | Habitantes por farmacia |
|-----------|---|------------|---------------------|------------|-------------------------|
| | Países | % sobre UE | Número | % sobre UE | |
| Sí | 15 | 55,56% | 99.636 | 64,03% | 2.864 |
| No | 12 | 44,44% | 55.983 | 35,97% | 3.712 |

Tabla 7. Comparativa del número de farmacias y habitantes por farmacia según los criterios de establecimiento

En la Tabla 8 advertimos que en el caso de los países donde los medicamentos sin receta sólo se venden en las oficinas de farmacia, el número de oficinas es superior en más de un 73% al de los países donde dicha restricción no se da. Además presentan una ratio de habitantes por farmacia inferior en 1.106 habitantes. En este caso, la proporción de países que presentan esta característica es superior a la proporción de los países que no la tienen. Representando, en este caso, la mayor diferencia de las tres variables analizadas. Además, estos países aglutinan la mayoría de las farmacias de la UE27, con un 63,45%.

| | Medicamentos sin receta sólo en farmacia | | Número | | Habitantes por farmacia |
|-----------|---|------------|--------|------------|----------------------------|
| | Países | % sobre UE | Número | % sobre UE | |
| Sí | 16 | 59,26% | 98.745 | 63,45% | 2.765 |
| No | 11 | 40,74% | 56.874 | 36,55% | 3.870 |

Tabla 8. Comparativa del número de farmacias y habitantes por farmacia según si los medicamentos sin receta sólo se venden en la farmacia

Como se aprecia en las tablas anteriores, en aquellos modelos donde las tres características están presentes, se presenta una menor ratio de habitantes por farmacia. Aparte de las diferencias comentadas en base a las tres variables analizadas, las principales diferencias estructurales se presentan en cuanto a la financiación de la asistencia farmacéutica, la cartera de servicios, la financiación de productos y la participación de los usuarios en el pago de medicamentos (14).

En la Figura 3 se observan los porcentajes de países de la UE27 en los que está presente un conjunto de características en su sistema farmacéutico, utilizados en el trabajo de Enrique Granda (14) sobre los modelos de farmacia en la UE27.

En dicha figura se ha incluido una marca del 50% para facilitar la interpretación de los datos.

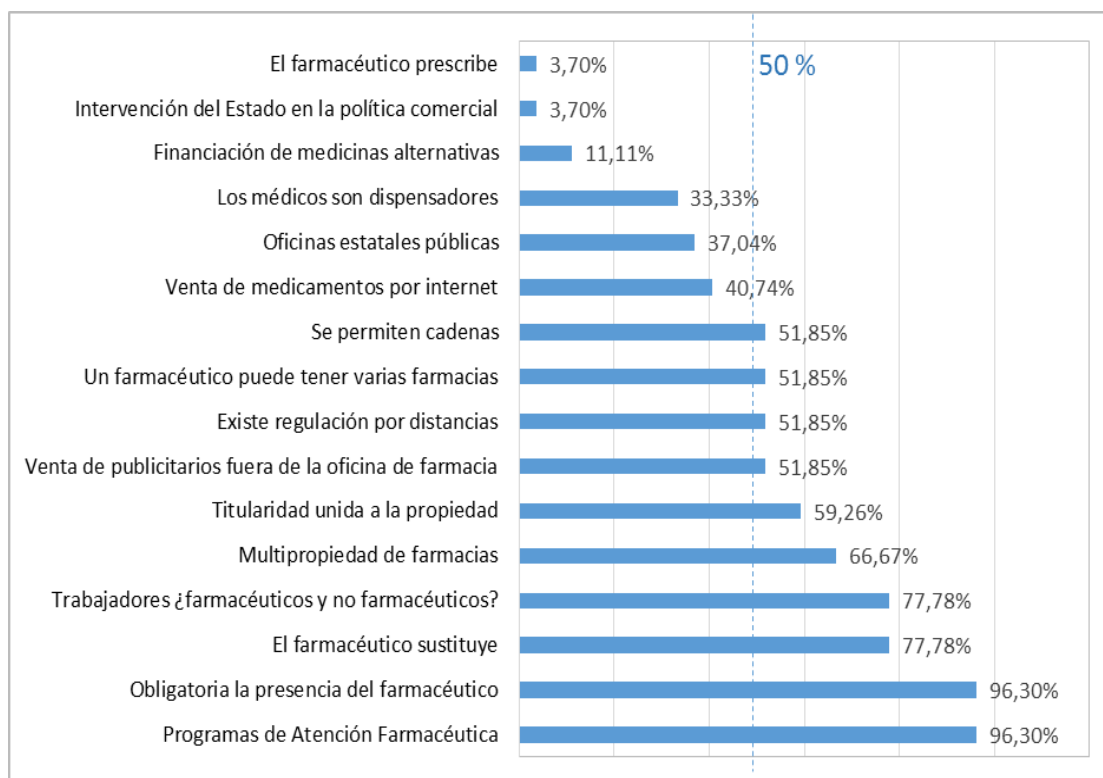


Figura 3. Porcentaje de Estados de la Unión Europea de los 27 que presentan la característica medida en su sistema farmacéutico

Podemos observar en esta figura como de los 16 aspectos considerados, 10 de ellos se presentan en más de la mitad de los países y sólo 6 se dan en menos de la mitad de los países de la UE27. También destaca, que en el 96,30% de los países, es obligatoria la presencia del farmacéutico en la oficina de farmacia. Ya vimos en la Tabla 6 que el binomio titularidad-propiedad se presenta en el 65,50% de los países de la UE27; es decir, que en prácticamente en todos los países hay un farmacéutico en la oficina de farmacia (propietario o no). De la misma forma, en el 96,30% de los países existen programas de Atención Farmacéutica en las oficinas de farmacia.

En el otro extremo, observamos la existencia de dos características que sólo se dan en un país. Concretamente, España es el único país en el que se da la intervención del Estado en la política comercial que limita los descuentos y la capacidad de gestión de las farmacias y el Reino Unido es el único en el que existe la capacidad de prescripción de los farmacéuticos (aunque no ejercida desde las oficinas de farmacia).

Cabe destacar también como en 11 países (44,44%) se permite la venta de medicamentos por internet (se prevé que este indicador aumentará en los próximos meses). En 9 países (33,33%) se permite dispensar a los médicos en determinadas zonas rurales y en 14 (51,85%) se da la existencia de cadenas, hecho que coincide con aquellos países en los que no existen criterios de establecimiento demográfico y/o geográfico como se apreció en la Tabla 6.

Respecto a la situación del pago de medicamentos en varios países de la UE podemos observar la siguiente tabla (14):

| País | Participación en el pago de medicamentos |
|-------------|---|
| Alemania | En medicamentos el copago es del 10% hasta un máximo de 10 €. |
| Austria | En farmacia se pagan 4,35 € por prescripción con excepciones como enfermedades infecciosas y personas sin recursos. |
| Bélgica | El copago de los medicamentos se divide en cinco categorías según la gravedad de la enfermedad. Son gratuitos los destinados a enfermedades graves y llegan a pagar el 80% de los medicamentos considerados de utilidad terapéutica baja. |
| Francia | Establece un reembolso diferenciado en tres categorías, con excepciones que afectan a las poblaciones más desfavorecidas. |
| Grecia | En los medicamentos, los pacientes participan en un 25% del coste excepto en algunos, indicados en enfermedades crónicas, en los que se reduce al 10%. |
| Holanda | En los medicamentos los pacientes sólo pagan la diferencia del precio de referencia financiado. |
| Italia | Existe un pago fijo por los medicamentos no considerados esenciales. |
| Portugal | En los medicamentos se distinguen tres tipos: los financiados al 100%, al 60% y al 30% |
| Reino Unido | Aunque hay variaciones en función del territorio, existe un copago que oscila entre 5 y 7 €. |

Tabla 9. Participación en el pago de medicamentos en países de la UE27

Como vimos en la Tabla 2 en España el conjunto de aportaciones no superaba el 5,8% antes del citado Decreto y tras él, la aportación del usuario pasó a ser proporcional al nivel de renta que posee, con topes máximos de aportación y colectivos exentos. Como se observa en la Tabla 9, la media europea en copago se encuentra en torno al 16,5% (14).

La evolución del mercado farmacéutico en Europa

Dado el carácter de esta investigación, se considera necesario en este punto analizar la evolución del mercado farmacéutico europeo. Para ello, nos basaremos en los datos publicados por la AESGP (*Association Européenne des Spécialités Pharmaceutiques Grand Public*) que se ofrecen en la Tabla 10 sobre las ventas realizadas en el periodo 2011-2013. Los datos se han ordenado de menor a mayor TVMA (Tasa de Variación Media Acumulativa).

| Mercado farmacéutico total | 2011 | 2012 | 2013 | TVMA | Gráfico evolutivo |
|----------------------------|----------|----------|------------|--------|-------------------|
| Grecia ** | 3.792,4 | 3.114,2 | 2.794,4 ▼ | -9,68% | |
| Holanda | 5.916,8 | 5.308,0 | 5.017,4 ▼ | -5,35% | |
| Portugal | 2.973,1 | 2.631,9 | 2.525,1 ▼ | -5,30% | |
| Hungría | 2.900,5 | 2.542,2 | 2.518,8 ▼ | -4,59% | |
| República Checa | 1.545,0 | 1.439,0 | 1.369,0 ▼ | -3,95% | |
| Turquía * | 5.972,6 | 5.673,2 | 5.518,9 ▼ | -2,60% | |
| España | 19.407,0 | 18.292,0 | 18.070,0 ▼ | -2,35% | |
| Italia | 18.860,0 | 17.796,0 | 17.793,8 ▼ | -1,92% | |
| Bélgica | 3.219,0 | 3.124,0 | 3.053,0 ▼ | -1,75% | |
| Eslovenia | 545,9 | 525,0 | 523,8 ▼ | -1,37% | |
| Francia | 28.849,0 | 28.829,0 | 27.837,0 ▼ | -1,18% | |
| Irlanda * | 1.956,3 | 2.015,3 | 1.930,8 ▼ | -0,44% | |
| Dinamarca | 2.123,9 | 2.160,8 | 2.115,3 ▼ | -0,14% | |
| Reino Unido * | 11.656,0 | 12.211,8 | 11.694,9 ▲ | 0,11% | |
| Suiza * | 3.226,8 | 3.343,3 | 3.250,8 ▲ | 0,25% | |
| Polonia * | 6.342,7 | 5.570,5 | 6.449,6 ▲ | 0,56% | |
| Suecia | 4.082,2 | 4.151,9 | 4.192,9 ▲ | 0,90% | |
| Eslovaquia * | 1.337,1 | 1.306,6 | 1.385,6 ▲ | 1,19% | |
| Noruega | 2.458,4 | 2.522,4 | 2.548,0 ▲ | 1,20% | |
| Austria | 4.313,4 | 4.551,8 | 4.490,0 ▲ | 1,35% | |
| Finlandia | 3.101,5 | 3.199,4 | 3.270,3 ▲ | 1,78% | |
| Alemania | 39.308,0 | 39.853,0 | 42.140,0 ▲ | 2,35% | |
| Bulgaria | 1.073,0 | 1.128,0 | 1.198,0 ▲ | 3,74% | |
| Rumanía * | 2.593,0 | 2.792,5 | 2.980,2 ▲ | 4,75% | |
| Ucrania | 2.083,0 | 2.634,0 | 2.869,0 ▲ | 11,26% | |

Tabla 10. Ventas totales en millones de euros. A precios PVP excepto los marcados con * (precio de venta del fabricante) y con ** (precio al por mayor)

Fuente: AESGP (Association Européenne des Spécialités Pharmaceutiques Grand Public)

Dada la heterogeneidad de los precios (venta al público, fabricante y al por mayor) hemos optado por calcular para cada país la TVMA (Tasa de Variación Media Acumulativa) de esta forma podemos medir la evolución del mercado abstrayéndonos del tipo de precio considerado. La expresión de la TVMA es la siguiente:

$$TVMA_0^t = \left(\sqrt[t-0]{\frac{X_t}{X_0}} \right) - 1$$

donde X_t es el valor de la magnitud en el instante t o periodo final y X_0 su valor en el instante o periodo inicial.

Al haber ordenado los valores de la tabla de menor a mayor TVMA, puede observarse más fácilmente el comportamiento de los países.

En primer lugar, destaca la mayor bajada sufrida por Grecia, seguida de Holanda, Portugal, Hungría y República Checa. Los casos de Grecia y Portugal quedan bien determinados por la grave situación de crisis que han experimentado en el periodo considerado. Se observa, excepto en el caso de Grecia, como sus modelos farmacéuticos expuestos en la Tabla 4, son del tipo liberal (Holanda, Hungría y República Checa tienen exactamente el mismo modelo).

En el otro extremo de la tabla, podemos observar a Ucrania con una gran subida, seguida de Rumania, Bulgaria y Alemania. Compartiendo modelo farmacéutico estas dos últimas.

A modo general, 12 países de los 25 estudiados han experimentado un aumento de su volumen en el mercado farmacéutico mientras que 13 han padecido un descenso del mismo.

En el caso de España, experimentó un descenso medio anual del 2,35%. Al igual que Grecia y Portugal, España ha sufrido una crisis muy importante en ese período, que puede distorsionar la explicación del mismo en los términos de esta investigación.

1.2.1.3 El modelo farmacéutico español

Del conjunto de los países de la Unión Europea, España es el que disfruta de una de las redes farmacéuticas más extensas y con mejor servicio. Una red que llega a cada rincón del territorio y que garantiza que todos los medicamentos estén a disposición de todos los ciudadanos en las mismas condiciones: igual proximidad, igual precio e igual atención profesional. Ventajas que ofrece el modelo farmacéutico mediterráneo. Un modelo que concibe las farmacias como establecimientos sanitarios de interés público, en el que la Administración los regula en beneficio de los ciudadanos, siendo el único punto de venta autorizado de medicamentos.

El modelo español de ordenación farmacéutica presenta las siguientes características que lo definen:

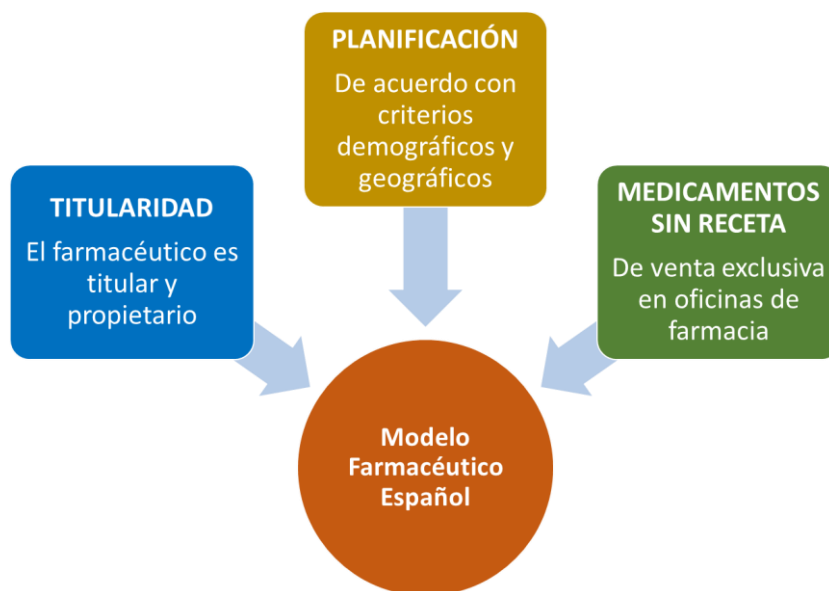


Figura 4. Características del modelo farmacéutico español

Como se recoge más adelante (15), el modelo español de ordenación farmacéutica permite que el 99% de la población española disponga de, al menos, una oficina de farmacia en el propio municipio, garantizando así el acceso en condiciones de igualdad al mismo medicamento con el mismo precio a lo largo de toda la geografía nacional.

Titularidad

En España, el titular de la oficina de farmacia ha de ser obligatoriamente un farmacéutico, solo o asociado con otros farmacéuticos, y cada farmacéutico sólo puede ser propietario de una oficina de farmacia. Se trata de un mecanismo que garantiza la independencia en las actuaciones de los farmacéuticos, al no estar condicionados por otros intereses que los estrictamente sanitarios. Además, evita conflictos de intereses con otros profesionales sanitarios prescriptores o con laboratorios farmacéuticos. Asegura también que los ciudadanos reciban asesoramientos independientes.

Esta característica del modelo farmacéutico español aporta los siguientes rasgos diferenciales (16):

- Servicio profesional.

Los ciudadanos pueden estar seguros de que siempre recibirán una atención profesional cuando acudan a su farmacia en busca de medicamentos, consejo o información. Farmacia que tiene la obligación de prestarles el servicio y donde hay un farmacéutico con capacidad legal necesaria para tomar rápidamente todas las decisiones relativas a las necesidades de medicación de los pacientes.

- Criterio sanitario.

Una atención basada en criterios sanitarios, que supone no sólo la dispensación de medicamentos, el consejo, la información y la promoción de la salud de los pacientes, sino también, y muy importante, garantizando la presencia y suministro de medicamentos adecuados a las necesidades reales de la zona a la que atienden, dejando los condicionantes de tipo económico en segundo lugar.

- Concurso público.

La posibilidad de obtener una nueva farmacia está al alcance de todos los farmacéuticos, ya que las autorizaciones se adjudican por concurso público. Un proceso de evaluación transparente, dirigido por la Administración Autonómica, donde se puntúan los méritos académicos y profesionales de todos los candidatos.

- Incentivo profesional.

Este sistema supone un gran incentivo para la profesión. Prueba de ello es que España es el país de la Unión Europea con más estudiantes y licenciados en farmacia por año.

Planificación farmacéutica

En cuanto a la planificación de oficinas de farmacia, el artículo 2.2 de la Ley 16/1997, de 25 de abril, sobre Regulación de Servicios de las Oficinas de Farmacia, afirma que éstos se establecerán de acuerdo con criterios demográficos y geográficos. Dichos criterios persiguen una regulación de las oficinas de farmacia en beneficio de la sociedad y, por tanto, de los pacientes y no de los profesionales.

Para garantizar una perfecta distribución y un fácil acceso de los pacientes, las 21.559 oficinas de farmacia (17) forman una red de establecimientos repartidos por todo el territorio nacional. En base a los criterios de planificación de la Ley estatal 16/1997 citada, desde la década de los noventa, las diferentes Comunidades Autónomas que componen el territorio español han establecido en su ámbito territorial los criterios de planificación de farmacias, en función de módulos geográficos, poblacionales y distancias, teniendo en cuenta las características propias de cada territorio.

De esta manera, se ha logrado una distribución homogénea de farmacias, situándose la ratio media de habitantes por farmacia entre las más bajas de Europa (18) (2.186 habitantes/farmacia a 31 de diciembre de 2.013), lo que permite que el 99% de la población disponga de una farmacia en su población de residencia, ya sea de tipo rural, urbano, semiurbano o turístico. Esta circunstancia hace que en muchas poblaciones el farmacéutico sea el único profesional sanitario presente.

Los criterios españoles de planificación farmacéutica permiten que todos los ciudadanos puedan acceder al medicamento en condiciones de igualdad efectiva y, por tanto, sin discriminaciones y con la misma calidad de servicio y suministro, en todo el Estado. Todos los ciudadanos pueden escoger libremente la oficina de farmacia.

Esta característica del modelo farmacéutico español aporta los siguientes rasgos diferenciales:

- Reparto según necesidades de la población.

En España, la ordenación territorial se traduce en dos requisitos básicos: la obligación de guardar distancias de al menos 250 metros entre dos oficinas de

farmacia y deben dar servicio a un módulo de población que cambia en función de la misma. Turística, rural, general, etc. De esta manera, se reparten de forma homogénea. Cubren grandes ciudades, población rural, población turista, atendiendo a las necesidades de todos los sectores de población. De esta forma, se consigue un reparto óptimo en las ciudades y en las zonas rurales.

- Medicamentos accesibles para todos.

Todos los ciudadanos pueden acceder a un bien de primera necesidad, como es el medicamento, de forma cómoda y rápida. A pocos metros de su casa y muy cerca también de los centros de salud públicos. Más calidad de vida que, en el caso de las personas dependientes, supone una mejora fundamental.

- Favorece a personas dependientes.

Precisamente este sector, objetivo prioritario de las políticas sociales europeas, es uno de los que más se beneficia del reparto planificado de las farmacias. Ancianos, discapacitados o enfermos crónicos pueden acudir por sus propios medios, ya sea desde su casa o desde su centro médico, al lugar donde se adquieren los tratamientos prescritos.

- Fijación de la población rural.

La planificación farmacéutica ayuda también a resolver otra de las preocupaciones actuales de la Unión Europea: la despoblación de las zonas rurales. La amplia red de farmacias en España lleva la atención sanitaria hasta los pueblos más apartados. De hecho, en muchos casos, es el único servicio del que dispone una población. Un servicio lo suficientemente importante como para evitar que los habitantes abandonen la zona.

- Atiende necesidades turísticas.

Un servicio que actúa, además, como reclamo turístico. El reparto de las farmacias presta especial atención a las necesidades de la población extranjera

que pasa sus vacaciones en España y que a menudo se instala en zonas apartadas de los grandes núcleos de población.

La Figura 5 muestra el mapa de cobertura de las oficinas de farmacia en España (19) como fiel reflejo de este criterio.

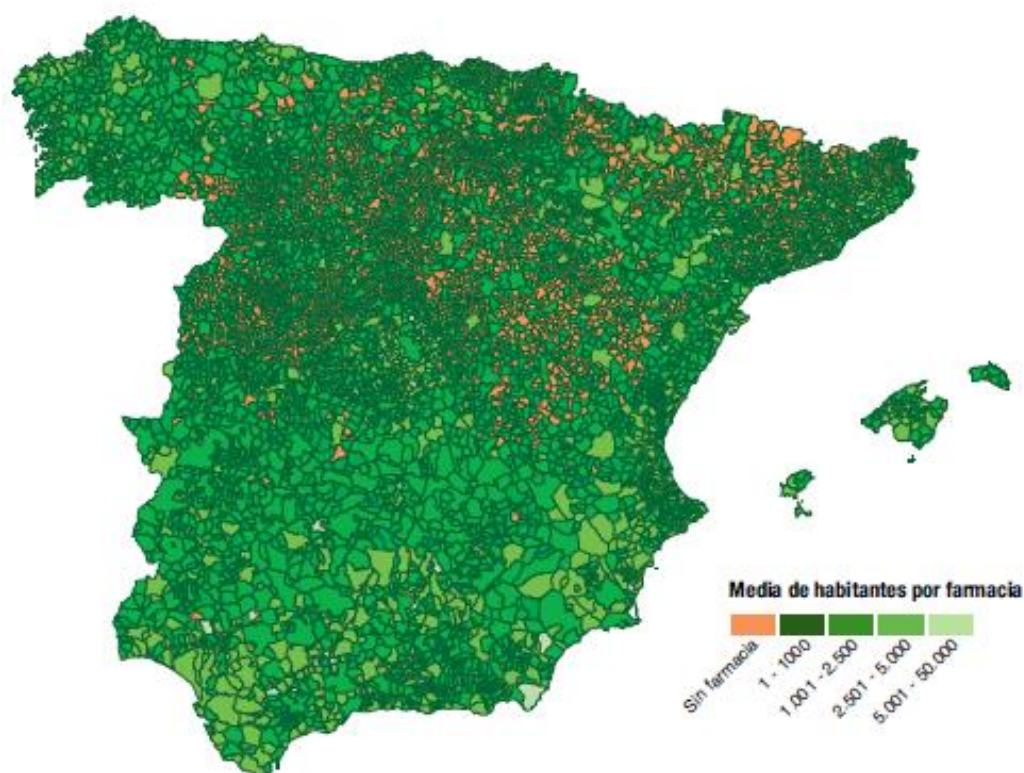


Figura 5. Mapa español de cobertura de oficinas de farmacia

Dispensación de medicamentos sin receta

El ordenamiento español, en su Ley 29/2006, de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios (20), reconoce expresamente la dispensación de los medicamentos sin prescripción en las oficinas de farmacia. El texto establece además que la venta de medicamentos sin receta por Internet se realizará desde una farmacia, con la intervención de un farmacéutico y con el asesoramiento personalizado previo, lo que sitúa a la normativa española entre las más restrictivas de Europa e introduce importantes cautelas en defensa del derecho de los ciudadanos a la protección de la salud.

Marco legal y funciones de la oficina de farmacia

Según aparece en el artículo 1 de la Ley 16/1997, de 25 de abril, de Regulación de Servicios de las Oficinas de Farmacia (21), la definición y funciones de las oficinas de farmacia es: en los términos recogidos en la Ley 14/1986, General de Sanidad, de 25 de abril, y la Ley 25/1990, de 20 de diciembre, del Medicamento, las **oficinas de farmacia son establecimientos sanitarios privados de interés público**, sujetos a la planificación sanitaria que establezcan las Comunidades Autónomas, en las que el farmacéutico titular-propietario de las mismas asistido, en su caso, de ayudantes o auxiliares, deberá prestar los siguientes servicios básicos a la población:

1. La adquisición, custodia, conservación y dispensación de los medicamentos y productos sanitarios.
2. La vigilancia, control y custodia de las recetas médicas dispensadas.
3. La garantía de la atención farmacéutica, en su zona farmacéutica, a los núcleos de población en los que no existan oficinas de farmacia.
4. La elaboración de fórmulas magistrales y preparados oficinales, en los casos y según los procedimientos y controles establecidos.
5. La información y el seguimiento de los tratamientos farmacológicos a los pacientes.
6. La colaboración en el control del uso individualizado de los medicamentos, a fin de detectar las reacciones adversas que puedan producirse y notificarlas a los organismos responsables de la farmacovigilancia.
7. La colaboración en los programas que promuevan las Administraciones sanitarias sobre garantía de calidad de la asistencia farmacéutica y de la atención sanitaria en general, promoción y protección de la salud, prevención de la enfermedad y educación sanitaria.
8. La colaboración con la Administración sanitaria en la formación e información dirigidas al resto de profesionales sanitarios y usuarios sobre el uso racional de los medicamentos y productos sanitarios.

9. La actuación coordinada con las estructuras asistenciales de los Servicios de Salud de las Comunidades Autónomas.

10. La colaboración en la docencia para la obtención del título de Licenciado en Farmacia, de acuerdo con lo previsto en las Directivas Comunitarias, y en la normativa estatal y de las Universidades por las que se establecen los correspondientes planes de estudio en cada una de ellas.

Con respecto a estas funciones reguladas podemos resaltar (22) que la oficina de farmacia no debe verse como un comercio, si no como un “establecimiento sanitario privado e interés público”. Ni el medicamento como un producto de consumo, sino como un bien sanitario que no va destinado al consumidor habitual, sino a un consumidor especial: el paciente.

1.2.2 Los canales de distribución y dispensación de medicamentos en España

Para conocer mejor el mercado farmacéutico español es necesario conocer sus agentes principales y sobre todo los canales de distribución. En España, las únicas entidades que legalmente pueden distribuir medicamentos son los almacenes mayoristas y los laboratorios farmacéuticos.

En la normativa vigente, Ley 29/2006, de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios, artículo 8.p) (20), se define la “distribución mayorista de medicamentos” como toda actividad que consista en obtener, almacenar, conservar, suministrar o exportar medicamentos, excluida la dispensación al público de los mismos; siendo el capítulo II, del título IV, artículos del 68 al 71bis), de esa misma ley, el que regula expresamente la distribución de medicamentos.

La organización actual de la distribución de medicamentos española se ha ido consolidando durante el último medio siglo. Un periodo en el que el sector, uno de los más dinámicos y competitivos de la economía mundial, ha experimentado una transformación radical que ha desplazado la elaboración de fármacos desde la práctica artesanal de las boticas hacia su producción industrial en presentaciones ya dispuestas para su consumo por el usuario final (23).

Para que los productos de la industria farmacéutica y lo más importante, sus beneficios para la salud, lleguen a los pacientes es necesario que existan los canales de distribución adecuados. En el caso de los medicamentos en España estos canales son: las empresas de distribución mayorista, la red de oficinas de farmacia y los servicios farmacéuticos hospitalarios y centros de salud de atención primaria (estos en menor medida).

Además de las actividades propias de cualquier sistema de distribución como son la compra, la venta, la financiación, la logística, etc., la distribución mayorista de medicamentos tiene un papel muy relevante dentro la cadena de valor del medicamento; asegurar el acceso a los medicamentos y productos farmacéuticos a toda la población.

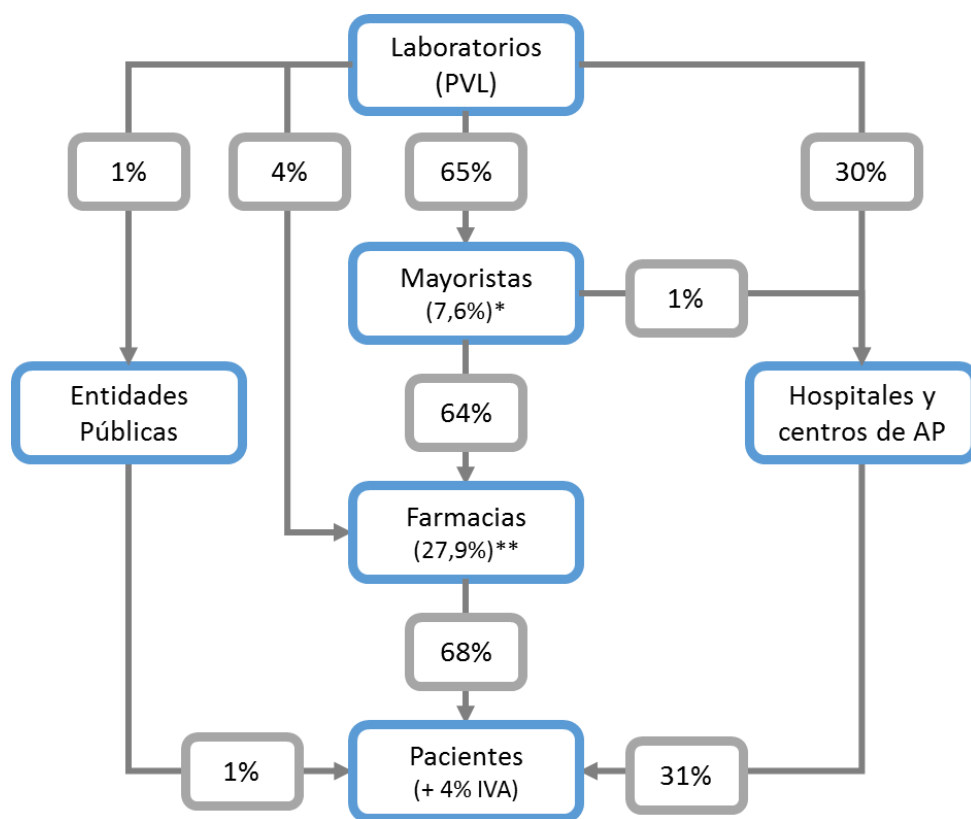


Figura 6. Canales de distribución y dispensación. Cuotas y márgenes

Fuente: Memoria farmaindustria 2012 y estimaciones IMS, FEDIFAR

* Margen correspondiente a la distribución de medicamentos (24) de uso humano fabricados industrialmente para las presentaciones de medicamentos cuyo precio de venta de laboratorio sea igual o inferior a 91,63€

** Margen correspondiente a las oficinas de farmacia (24) por dispensación y venta al público de medicamentos de uso humano fabricados industrialmente, para aquellos medicamentos cuyo precio industrial sea igual o inferior a 91,63 €

Según observamos en la Figura 6, el 94,1% de los medicamentos y productos farmacéuticos distribuidos por las oficinas de farmacia es proporcionado por las distribuidoras farmacéuticas. En España (25), el sector de la distribución mayorista se compone de 52 empresas, de las que los 6 principales grupos representan el 75,5% del mercado, por lo tanto, se trata de un sector muy concentrado en España.

Como comentábamos antes, en los últimos años la industrialización en el sector ha generado tres grandes grupos en el mercado farmacéutico; por un lado, la fabricación de medicamentos para su consumo por el usuario final (los laboratorios), por otro la distribución (mayoristas) y por último, la dispensación (las oficinas de farmacia). Sus cuotas de mercado las podemos observar en la Figura 6.

La distribución mayorista española se clasifica de la siguiente forma (25):

- Almacenes de capital no farmacéutico:
Son sociedades anónimas con o sin participación de capital extranjero.
- Almacenes de capital farmacéutico:
 - Cooperativas (sociedades de capital farmacéutico cuyos dueños son las propias farmacias).
 - Centros (sociedades anónimas).

| Tipología | Cuota de Mercado | Número de empresas |
|--|------------------|--------------------|
| Cooperativas | 75,20% | 30 |
| Sociedades mercantiles de otra tipología | 18,80% | 18 |
| Centros Farmacéuticos (S.A. con farmacéuticos titulares de farmacias como accionistas) | 6,00% | 4 |
| Total | 100,00% | 52 |

Tabla 11. Tipo y peso de la distribución mayorista de gama completa

Fuente: FEDIFAR (Federación de Distribuidores Farmacéuticos). Octubre 2013

Como se observa en la Tabla 11, una característica importante de la distribución mayorista española es que en su mayor parte está manos de colectivos boticarios.

Con todo y con ello, el número de mayoristas en España es sustancialmente mayor al de otros países de su entorno como Bélgica, Francia, Reino Unido o Alemania (26).

Dado que la farmacia hospitalaria es otro de los agentes consumidores de medicamentos (como veremos más adelante, el gasto farmacéutico hospitalario supone más del 33% del gasto farmacéutico público en España), en los párrafos siguientes describimos sus principales características.

Los servicios de farmacia hospitalaria europeos han desarrollado diferentes iniciativas que contribuyen a la utilización segura, efectiva y económica de los medicamentos (27), integrando funciones de preparación, compra, almacenamiento, distribución y dispensación propias de todos los demás actores de la cadena del medicamento.

Las funciones de los servicios de farmacia hospitalaria van más allá de la mera dispensación ya que ésta sólo supone una de sus actividades.

Las principales funciones de la farmacia hospitalaria española en cuanto a selección de medicamentos se pueden resumir en las siguientes líneas:

1. Creación de herramientas metodológicas y estandarización de procedimientos en evaluación y selección de medicamentos.
2. Elaboración de informes técnicos de apoyo a la decisión de selección de medicamentos con criterios transparentes de evaluación de eficacia, seguridad, coste-efectividad e impacto presupuestario.
3. Formación e investigación en selección de medicamentos.
4. Colaboración con otros profesionales en la toma de decisiones de selección y posicionamiento de los fármacos aportando una visión complementaria en el ejercicio de la selección, formando parte de los equipos multidisciplinares para esta función, las Comisiones de Farmacia y Terapéutica.

Dentro de la dispensación de un hospital se pueden diferenciar dos casos (28):

1. Atención farmacéutica intrahospitalaria: dispensación de medicamentos a pacientes ingresados o que acuden al hospital para recibir su tratamiento.
2. Atención farmacéutica extra-hospitalaria: dispensación de medicamentos a pacientes para su administración fuera del hospital (alta hospitalaria y atención domiciliaria al paciente).

El volumen que representa la medicación intrahospitalaria en el conjunto de la dispensación del hospital llega a ser de dos tercios frente a la dispensación extra-hospitalaria (29).

Composición relativa del precio (PVP) de los medicamentos financiados por el SNS y participación en el mismo de cada agente del sector

El precio final de los medicamentos está compuesto por los siguientes componentes (30):

- Precio Venta Laboratorio (PVL).
- Margen del mayorista.
- Margen de la farmacia.
- Impuesto sobre el valor añadido (IVA).

La estructura del precio de los medicamentos en Europa tiene unas diferencias bastante notables en cada una de las variables que componen el precio final de los medicamentos.

Respecto al margen comercial, la distribución mayorista tiene, al igual que las oficinas de farmacia, un sistema de retribución basado en un margen o porcentaje, fijado por ley sobre el precio del producto, que podemos observar entre paréntesis en la Figura 6.

En la Tabla 12, se detalla la composición del PVP de los medicamentos financiados por el Sistema Nacional de Salud (31).

| Medicamentos en función del margen de la farmacia | Laboratorio (% sobre el total del PVP) | Margen Distribución Mayorista (% sobre el total del PVP) | Margen Farmacia (% sobre el total del PVP) | PVP (BASE = 100) |
|--|---|---|--|---------------------|
| Medicamentos en general | 66,62 | 5,48 | 27,9 | 100 |
| Medicamentos de margen fijo por envase | $\geq 66,62$ | $\leq 5,48$ | $\leq 27,9$ | 100 |
| Precio medio de los medicamentos financiados | 67,94 | 5,26 | 26,8 | 100 |
| (Menos) Aportaciones y Deducciones | -3,74 | -0,29 | -3,9 | -7,93 |
| Participación en el PVP de los medicamentos financiados por el Sistema Nacional de Salud | 64,2 | 4,97 | 22,9 | 92,07 |

Tabla 12. Composición relativa del precio (PVP) de los medicamentos financiados por el SNS y participación en el mismo de cada agente del sector

Fuente: Medicamentos y Farmacia en Cifras 2013. Consejo General de Colegios Farmacéuticos

A continuación, se ofrece la evolución y estado actual de las aportaciones y deducciones en España:

- Desde el año 2000, la farmacia realiza aportaciones directas al Sistema Nacional de Salud por la aplicación de la escala de deducciones establecida en el RDL 5/2000 sobre la facturación de medicamentos de cada farmacia (32).
- Desde el año 2010, por la aplicación del RDL 8/2010 (33), se aplican deducciones del 7,5% y del 4% sobre el precio de los medicamentos financiados con cargo a fondos públicos excluidos del sistema de precios de referencia, deducciones que se reparten entre todos los agentes del sector (Industria, Distribución y Farmacia).
- Mediante el RDL 9/2011 (34), se introdujo en el RDL 8/2010 una nueva deducción del 15% sobre los medicamentos para los que hayan transcurrido diez años

desde la fecha en que se hubiese adoptado la decisión de financiar con fondos públicos, o del 11% en el caso de haber sido autorizada una nueva indicación.

Desde el 1 de julio de 2010, los márgenes de los medicamentos son los siguientes:

| Rango según el PVL | Margen Distribución | Margen Farmacia |
|---|--|---|
| Medicamentos con precio Laboratorio menor o igual a 91,63 euros | 7,6 % PVD (Precio de Venta del Distribuidor) | 27,9% PVP Precio de Venta al Público) |
| Medicamentos con precio Laboratorio superior a 91,63 euros e igual o inferior a 200 euros | 7,54 euros por envase | 38,37 euros por envase |
| Medicamentos con precio Laboratorio superior a 200 euros e igual o inferior a 500 euros | 7,54 euros por envase | 43,37 euros por envase |
| Medicamentos con precio Laboratorio superior a 500 euros | 7,54 euros por envase | 48,37 euros por envase |

Tabla 13. Márgenes de los medicamentos de distribución mayorista y farmacias

1.2.3 La oficina de farmacia española en cifras

Pasamos a describir con más detalle la caracterización de la oficina de farmacia dada su relevancia en esta investigación. Por ello, se dedica este apartado específico.

Según el Consejo General de Colegios Oficiales de Farmacéuticos (31) los datos que siguen a continuación, reflejan la realidad de la oficina de farmacia española (datos a fecha 5 de febrero de 2015):

- *La Farmacia española, respecto a los países de nuestro entorno, es la más próxima y accesible al ciudadano. El 99% de la población dispone de oficina de farmacia en su propio municipio.*
- *Aproximadamente 1 de cada 4 oficinas de farmacia están ubicadas en el medio rural y, concretamente, 3 de cada 4 farmacias rurales, en municipios de menos de 5.000 habitantes.*
- *España es uno de los países europeos que más farmacias tiene por habitante. Precisamente, este promedio se sitúa actualmente en un nivel crítico, que, de disminuir, incidiría notablemente en la calidad del servicio. En España el número de habitantes por oficina de farmacia se sitúa en torno a los 2.200, frente a países como Suecia con más de 10.000 habitantes por oficina de farmacia.*
- *Los datos procedentes de Farmaindustria, la OCDE y la Agrupación Farmacéutica de la Unión Europea reflejan que España, respecto a los países que integran la Unión Europea, cuenta con:*
 - *El precio medio de los medicamentos más bajo, sin tener en cuenta a Francia.*
 - *Las ventas medias por oficina de farmacia es una de las más bajas de Europa.*
- *Cada día están de servicio de guardia en España, durante las 24 horas, un número de farmacias suficiente para atender las urgencias de toda la población española.*
- *El actual ordenamiento farmacéutico permite que el medicamento llegue en condiciones de igualdad al ciudadano del medio urbano y del pueblo más recóndito de la geografía española.*

- *El gasto en medicamentos permite que enfermedades que antes eran de tratamiento hospitalario hoy lo sean de ambulatorio, con la consiguiente amortización de camas hospitalarias.*
- *La Farmacia es uno de los servicios mejor valorados por la sociedad. Las reclamaciones sobre la Farmacia que se reciben en las organizaciones de consumidores representan sólo el 0,27%, mientras que otros servicios llegan al 14%.*
- *Cada año las farmacias ofrecen 182 millones de consejos sanitarios ajenos a la dispensación. Así, uno de cada 3 españoles que entra en una farmacia no adquirió un medicamento y recibió a cambio un consejo sanitario.*
- *El valor de los actos equivalentes, medida del ahorro producido al Sistema Sanitario, alcanza un nivel anual que supera los 1.700 millones de euros.*

A continuación, y haciendo uso de las estadísticas publicadas por el Consejo General de Colegios Oficiales de Farmacéuticos (17) , pasamos a caracterizar la oficina de farmacia española.

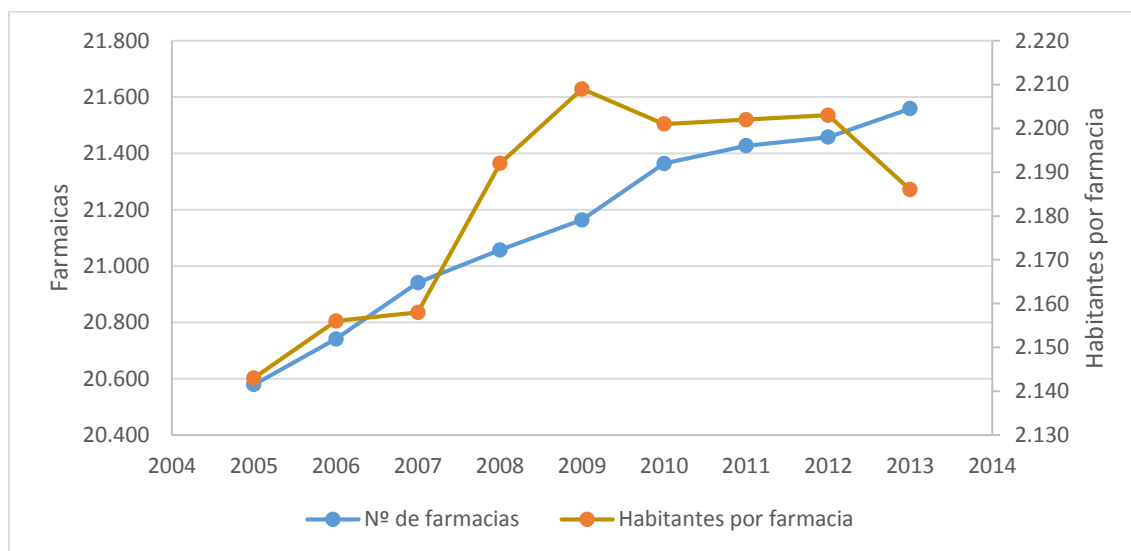


Figura 7. Evolución del número de farmacias y del ratio habitantes/farmacia en España (2005-2013)

Fuente: Medicamentos y Farmacia en Cifras 2013. Consejo General de Colegios Farmacéuticos y elaboración propia

Como vemos en la Figura 7, en el período 2005-2013 el número de farmacias en España presenta una tendencia ascendente y continua con una tasa de variación media

acumulada del 0,58% pasando de las 20.579 farmacias a las 21.559. Esto supone un incremento neto de 980 farmacias en el periodo.

Respecto al ratio de habitantes/farmacia, aunque presenta una tendencia creciente, su comportamiento ha sido irregular. Dado que el número de farmacias ha tenido un crecimiento constante, la explicación de las oscilaciones de este indicador viene dada por la variación de la población española en este periodo (Figura 8) cuya tasa de variación media acumulada (TVMA) ha sido del 0,95% lo que supone un crecimiento medio anual del 64% mayor al del número de oficinas de farmacia.

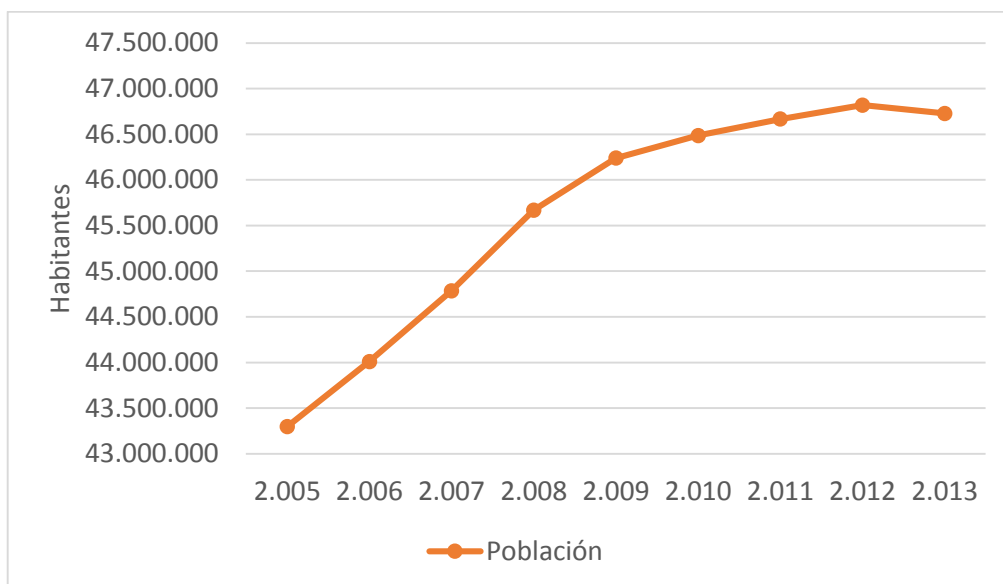


Figura 8. Evolución de la población española según padrón en el periodo (2003-2013)

Fuente: Instituto Nacional de Estadística y elaboración propia

A nivel de comunidades autónomas, destacan como las que tienen un mayor número de farmacias Andalucía, Cataluña y Madrid y con menor número de farmacias La Rioja, Melilla y Ceuta (véase Tabla 14).

Respecto a la ratio de habitantes/farmacia, cabe destacar que las que menor ratio presentan son Navarra, Castilla y León, y Extremadura, mientras que las de mayor ratio son Melilla, Ceuta y Canarias. Hay 7 comunidades por debajo de la ratio media nacional y 12 por encima.

| Autonomía | Farmacias | Habitantes por Farmacia |
|----------------------|---------------|-------------------------------|
| Andalucía | 3.599 | 2.345 |
| Aragón | 731 | 1.843 |
| Asturias | 456 | 2.342 |
| Baleares | 436 | 2.550 |
| Canarias | 706 | 3.001 |
| Cantabria | 254 | 2.330 |
| Castilla la Mancha | 1.274 | 1.649 |
| Castilla y León | 1.639 | 1.537 |
| Cataluña | 3.117 | 2.423 |
| Extremadura | 675 | 1.636 |
| Galicia | 1.345 | 2.056 |
| La Rioja | 156 | 2.064 |
| Madrid | 2.830 | 2.295 |
| Murcia | 566 | 2.601 |
| Navarra | 600 | 1.074 |
| País Vasco | 826 | 2.653 |
| Comunidad Valenciana | 2.303 | 2.221 |
| Ceuta | 24 | 3.508 |
| Melilla | 22 | 3.804 |
| ESPAÑA | 21.559 | 2.186 |

Tabla 14. Número de farmacias y ratio de habitantes/farmacia por comunidades autónomas

Fuente: Medicamentos y Farmacia en Cifras 2013. Consejo General de Colegios Farmacéuticos

En la Tabla 15 aparecen un conjunto de indicadores, que vamos a desgranar en los siguientes párrafos con la ayuda de gráficos y tasas de variación media acumulada (TVMA) que permitan una mejor interpretación de los mismos.

| Indicador | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|--|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Colegiados | 59.638 | 61.084 | 61.919 | 62.507 | 63.377 | 64.203 | 64.977 | 65.472 | 66.657 |
| % Mujeres colegiadas | 68,70% | 69,80% | 70,20% | 70,30% | 70,40% | 70,60% | 71,00% | 71,00% | 71,20% |
| % Hombres colegiados | 31,30% | 30,20% | 29,80% | 29,70% | 29,60% | 29,40% | 29,00% | 29,00% | 28,80% |
| Titulares de farmacia | 20.579 | 23.692 | 23.659 | 23.993 | 23.945 | 24.042 | 24.130 | 24.259 | 24.367 |
| % Mujeres titulares de farmacia | 60,60% | 62,80% | 63,20% | 63,30% | 63,00% | 63,00% | 63,50% | 63,50% | 63,70% |
| % Hombres titulares de farmacia | 39,40% | 37,20% | 36,80% | 36,70% | 37,00% | 37,00% | 36,50% | 36,50% | 36,30% |
| % Titulares Farmacia/Colegiados | 34,50% | 38,80% | 38,20% | 38,40% | 37,80% | 37,40% | 37,10% | 37,10% | 36,60% |

Tabla 15. Evolución del número de colegiados y titulares de farmacia por sexo en España en el período (2005-2013)

Fuente: Medicamentos y Farmacia en Cifras 2013. Consejo General de Colegios Farmacéuticos

Nota: hasta el año 2005 se consideraba un único titular por farmacia. Desde el año 2006, se incluyen a los copropietarios

Respecto al número de colegiados, Figura 9, se observa en el periodo estudiado un incremento neto de 7.019 colegiados, lo que supone un crecimiento medio anual acumulativo (TVMA) del 1,40%. En el caso del número de titulares de farmacia se ha producido un incremento neto de 3.788 titulares que ha supuesto un crecimiento medio anual acumulativo del 2,13%. No podemos obviar el hecho de que a partir del 2006 se contabilizan los copropietarios en el indicador número de titulares. Por lo tanto si no contamos con el 2005, el crecimiento medio anual acumulativo sería del 0,40%. Bastante menor que el experimentado por el del número de colegiados.

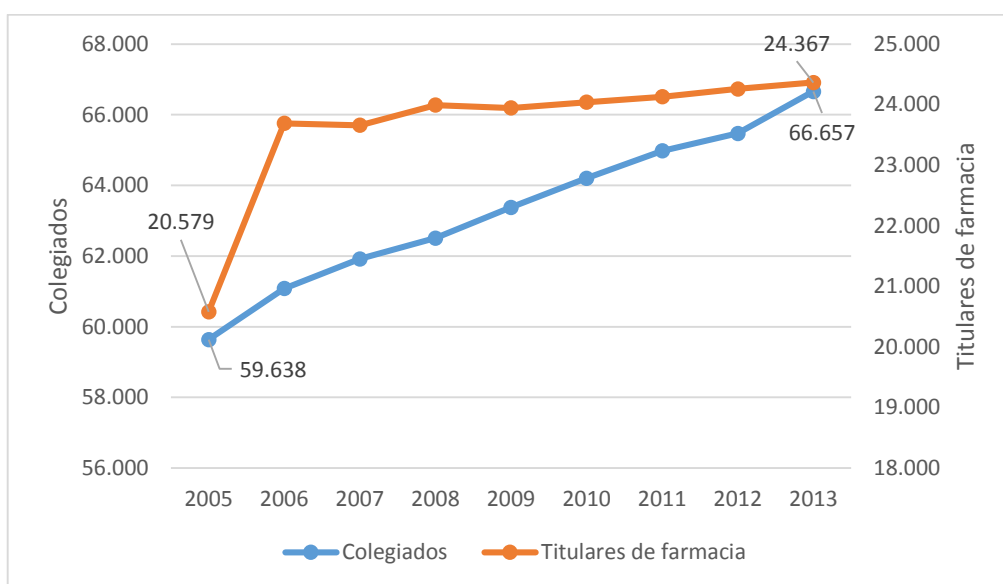


Figura 9. Evolución del número de colegiados y titulares de farmacia en España en el período (2005-2013)

Fuente: Medicamentos y Farmacia en Cifras 2013. Consejo General de Colegios Farmacéuticos y elaboración propia

En lo que se refiere al sexo del colegiado de farmacia, se puede observar que en actualidad el 71,20% de los colegiados son mujeres frente a un 28,80%, que son hombres. Esta diferencia ha ido en aumento en el periodo seleccionado como puede verse en la Figura 10.

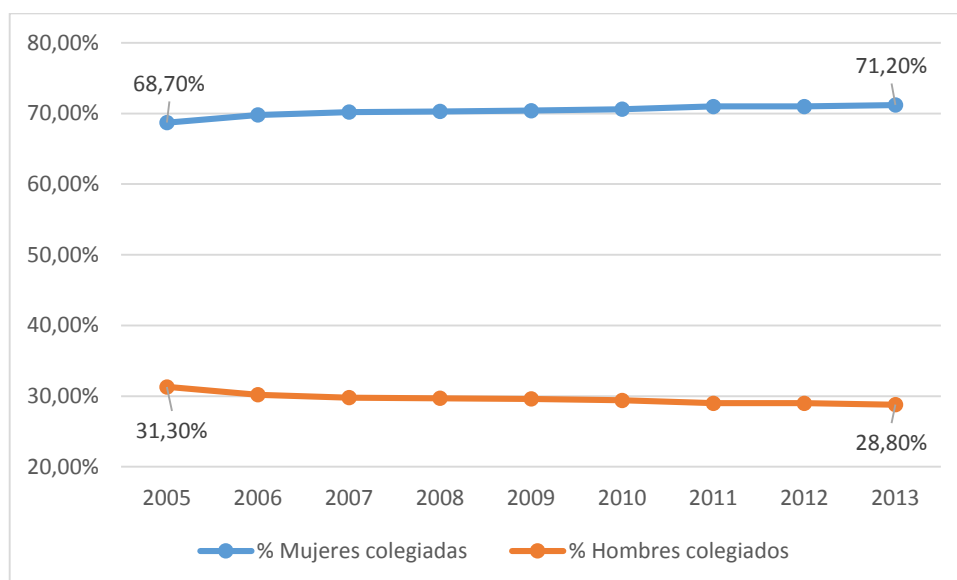


Figura 10. Evolución del número de colegiados según sexo en España en el período (2005-2013)

Fuente: Medicamentos y Farmacia en Cifras 2013. Consejo General de Colegios Farmacéuticos y elaboración propia

Respecto al sexo de los titulares de farmacia, como se aprecia en la Figura 11, las mujeres son mayoría. En este caso las diferencias se atenúan algo pero aun así, la diferencia es elevada entre ambos sexos; el 63,70% de los titulares de farmacia son mujeres.

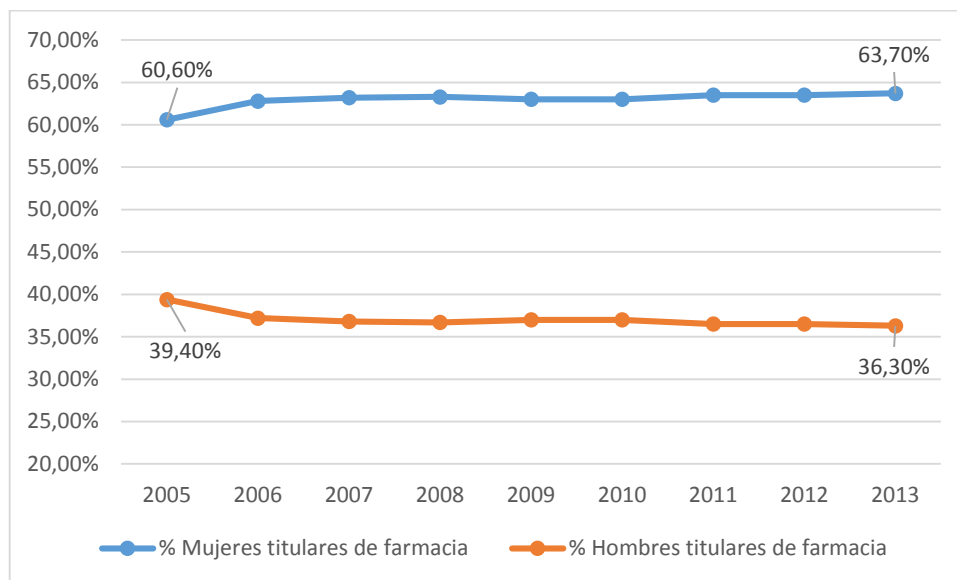


Figura 11. Evolución del número de titulares según sexo en España en el período (2005-2013)

Fuente: Medicamentos y Farmacia en Cifras 2013. Consejo General de Colegios Farmacéuticos y elaboración propia

Por último, observando la Figura 12, vemos que el comportamiento de la ratio titulares/colegiados en el período 2005-2013, ha presentado tendencia decreciente. Si tenemos en cuenta el hecho de que en 2006 se incorporan los copropietarios al número de titulares y consideramos, por tanto, el período 2006-2013, observamos un decrecimiento medio anual acumulativo del 0,83% (TVMA). En el año 2013, último valor publicado hasta la fecha, el 36,60% de los colegiados en España son titulares de una farmacia.

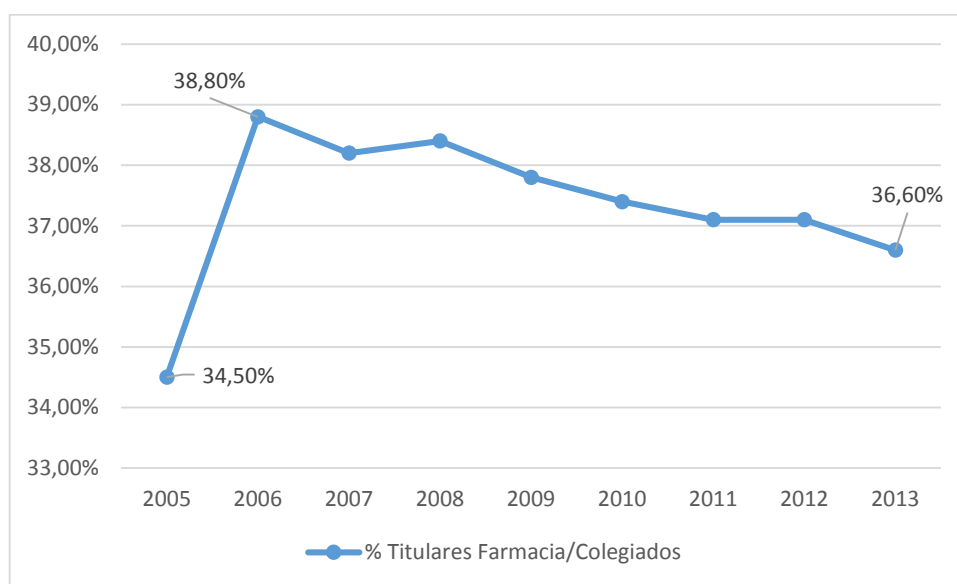


Figura 12. Evolución del ratio Titulares/Colegiados en España en el período (2005-2013)

Fuente: Medicamentos y Farmacia en Cifras 2013. Consejo General de Colegios Farmacéuticos y elaboración propia

1.2.4 El gasto sanitario en España

El sector de la salud forma parte esencial de la economía productiva y social de los países avanzados y es generador continuo de riqueza, prosperidad y empleabilidad cualificada. Su alta interdependencia con otros mercados estratégicos y globales, convierten a su cadena de diseño y de valor en un clúster de innovación disruptiva de primer orden. En los últimos 40 años esta economía del bienestar ha sido el sector con más rápido crecimiento en Europa y Estados Unidos (1) .

Dado que el mercado farmacéutico se enmarca dentro del mercado sanitario, y por ello condiciona al farmacéutico, en primer lugar vamos a analizar sucintamente la evolución del gasto sanitario en España con el objeto de mejorar el análisis del comportamiento del mercado farmacéutico.

En nuestro caso, haremos referencia al período 2008-2012 puesto que es el último periodo disponible como estadística oficial cuya publicación se denomina “Estadística de Gasto Sanitario Público” del Ministerio de Sanidad, Servicios Sociales e Igualdad (35).

| | Gasto Sanitario Público | % sobre PIB | € por habitante |
|-------------|--|------------------------|----------------------------|
| 2008 | 66.918,00 | 6,20% | 1.450 |
| 2009 | 70.579,00 | 6,70% | 1.510 |
| 2010 | 69.357,00 | 6,60% | 1.475 |
| 2011 | 68.050,00 | 6,50% | 1.442 |
| 2012 | 64.150,00 | 6,20% | 1.357 |
| TVMA | -1,05% | | -1,64% |

Tabla 16. Gasto sanitario público consolidado. Millones de euros, porcentaje sobre PIB y euros por habitante. España, 2008-2012

Fuente: Estadística de Gasto Sanitario Público 2012: Principales resultados. Ministerio de Sanidad, Servicios Sociales e Igualdad

Como se puede observar en la Tabla 16 el gasto sanitario público ha decrecido en 2.768 millones de euros, lo que equivale a un 4,13% respecto al 2008. Considerando el quinquenio 2008-2012, la tasa de variación media acumulada se sitúa en el -1,05% lo que implica que el gasto sanitario público ha decrecido a razón de un 1,05% al año de forma acumulada en el periodo 2008-2012. El gasto por habitante también ha decrecido en dicho periodo a razón de un 1,64% al año de forma acumulada.

Por otro lado, el PIB se redujo en este periodo y en términos medios anuales, un 1,4%.

En los estudios e informes del sector, es común encontrar los indicadores “porcentaje del gasto sanitario público sobre el PIB” y “porcentaje del gasto farmacéutico público sobre el PIB” con objeto de valorar la capacidad de la economía para hacer frente al gasto sanitario y a la factura en medicamentos.

Ahora analizaremos estos valores según la clasificación funcional del gasto sanitario público. En dicha clasificación, es muy importante resaltar que la función de Farmacia recoge el gasto por las recetas médicas expedidas, el reintegro de gastos farmacéuticos y el gasto en otros productos de suministro directo a los hogares. No se incluye el gasto en productos farmacéuticos que puedan dispensarse ambulatoriamente en los hospitales (programas selectivos de revisión de medicamentos y medicación de urgencia).

| | Servicios hospitalarios y especializados | Servicios primarios de salud | Servicios de salud pública | Servicios colectivos de salud | Farmacia | Traslado, prótesis y aparatos terapéuticos | Gasto de capital | Total consolidado |
|-------------|--|------------------------------------|----------------------------------|-------------------------------------|---------------|---|---------------------|----------------------|
| 2008 | 37.018 | 10.396 | 784 | 1.907 | 12.690 | 1.230 | 2.833 | 66.858 |
| 2009 | 39.251 | 10.775 | 1.158 | 2.074 | 13.415 | 1.351 | 2.556 | 70.580 |
| 2010 | 38.908 | 10.610 | 761 | 2.058 | 13.407 | 1.467 | 2.205 | 69.416 |
| 2011 | 39.803 | 10.182 | 750 | 2.088 | 12.334 | 1.393 | 1.500 | 68.050 |
| 2012 | 38.908 | 9.505 | 699 | 2.012 | 10.708 | 1.248 | 1.071 | 64.151 |
| TVMA | 1,25% | -2,22% | -2,83% | 1,35% | -4,16% | 0,36% | -21,59% | -1,03% |

Tabla 17. Gasto sanitario público consolidado según clasificación funcional. Millones de euros. España, 2008-2012

Fuente: Estadística de Gasto Sanitario Público 2012: Principales resultados. Ministerio de Sanidad, Servicios Sociales e Igualdad

Si prestamos atención a las Tasas de Variación Media Acumulada (TVMA) de la Tabla 17 podemos observar que la mayor bajada se produce en Farmacia con un descenso medio anual acumulado del 4,16% que suponen 1.982 millones de € en el periodo 2008-2012.

Como vemos en la Tabla 18, la función que mayor peso tiene en el gasto de la asistencia sanitaria pública es la de los servicios hospitalarios y especializados, que representaron un 60,65% del gasto total consolidado en el año 2012. En segundo lugar, se encuentra el gasto en farmacia, con un 16,69%, y en tercer lugar los servicios primarios de salud con un 14,82% como se detalla en dicha tabla.

En la serie del gasto en servicios de salud pública es importante señalar que el cambio brusco que se observa en el año 2009 obedece a la compra centralizada en el otoño de dicho año por el entonces Ministerio de Sanidad y Política Social de la vacuna contra la gripe A (H1N1), por un importe de 333 millones de euros (35).

| | Servicios hospitalarios y especializados | Servicios primarios de salud | Servicios de salud pública | Servicios colectivos de salud | Farmacia | Traslado, prótesis y aparatos terapéuticos | Gasto de capital | Total consolidado |
|--|--|------------------------------------|----------------------------------|-------------------------------------|---------------|---|---------------------|----------------------|
| 2008 | 55,37% | 15,55% | 1,17% | 2,85% | 18,98% | 1,84% | 4,24% | 100,00% |
| 2009 | 55,61% | 15,27% | 1,64% | 2,94% | 19,01% | 1,91% | 3,62% | 100,00% |
| 2010 | 56,05% | 15,28% | 1,10% | 2,96% | 19,31% | 2,11% | 3,18% | 100,00% |
| 2011 | 58,49% | 14,96% | 1,10% | 3,07% | 18,12% | 2,05% | 2,20% | 100,00% |
| 2012 | 60,65% | 14,82% | 1,09% | 3,14% | 16,69% | 1,95% | 1,67% | 100,00% |
| % promedio sobre el total | 57,18% | 15,18% | 1,22% | 2,99% | 18,45% | 1,97% | 3,00% | 100,00% |

Tabla 18. Cuota sobre el gasto sanitario público consolidado según clasificación funcional. España, 2008-2012

Fuente: Estadística de Gasto Sanitario Público 2012: Principales resultados. Ministerio de Sanidad, Servicios Sociales e Igualdad

En la Tabla 18 podemos observar como el gasto en farmacia ha perdido un peso de 2,29 puntos en el gasto sanitario público en el 2012 con respecto al 2008; esto supone la mayor pérdida experimentada por todas las funciones salvo en el caso del gasto en capital con una pérdida de 2,77 puntos (que tiene una incidencia marginal en el gasto sanitario público). El mayor peso ganado ha sido protagonizado por los servicios hospitalarios y especializados. Que incluso han experimentado un crecimiento medio en el periodo, a razón de un 1,25% anual acumulado.

A modo de resumen, podemos resaltar que el componente del gasto sanitario que más ha decrecido ha sido el farmacéutico y cuya cuota dentro del total del gasto también ha disminuido en más de dos puntos porcentuales.

1.2.5 El mercado farmacéutico español

El mercado del medicamento presenta diferencias sustanciales con respecto a otro tipo de mercados en los que se comercializan otro tipo de bienes y servicios. Estas diferencias serán tenidas en cuenta en esta investigación tanto para su justificación como para una mejor comprensión de los resultados.

Este apartado está dividido en dos grandes mercados desde el punto de vista de la oficina de farmacia: el mercado de los medicamentos de prescripción y el mercado del autocuidado de la salud.

El conocimiento de las características específicas del mercado es indispensable para la realización de las actividades enmarcadas en esta investigación.

En este sentido, y según los informes sectoriales y la literatura en este ámbito, es práctica del sector considerar las divisiones del mercado que aparecen en la Figura 13 que se recoge a continuación:

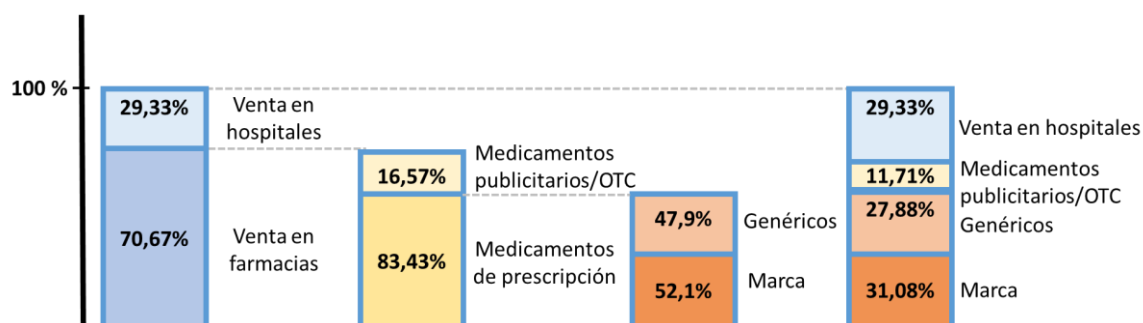


Figura 13. Canales, tipos de medicamentos y sus cuotas en el mercado farmacéutico según precio

Fuente: IMS Health- Análisis de la evolución del gasto farmacéutico público en España. Datos: 2014

A continuación vamos a detallar las definiciones y características principales de cada uno de los tipos de medicamentos que dan lugar a mercados con entidad propia dentro del mercado farmacéutico español.

1.2.5.1 El mercado de los medicamentos de prescripción médica

Este mercado está constituido por todos los medicamentos que el cliente o paciente puede adquirir en una oficina de farmacia siempre y cuando presente la prescripción correspondiente extendida por un médico, bien sea pública o privada.

Dentro de este mercado podemos distinguir dos tipos de productos (36):

- Productos con marca: sólo productos innovadores (nueva molécula obtenida a través de un proceso de investigación y desarrollo) y sus licencias, comercializados bajo el nombre de una marca comercial. En España, en estos momentos, es el grupo más importante en cuanto al volumen de ventas.
- Productos genéricos: productos con el mismo principio activo, la misma dosis, forma farmacéutica y características farmacocinéticas y farmacodinámicas que el producto original, con el que deben demostrar bioequivalencia terapéutica. Se nombran por la denominación común internacional (D.C.I.) del principio activo que contiene y el laboratorio comercializador. El uso racional del medicamento promulgado por la Organización Mundial de la Salud aboga por que “los pacientes reciban medicamentos adecuados a sus necesidades clínicas, a las dosis precisas, durante el periodo de tiempo adecuado y al menor coste posible para ellos y para la comunidad”. En este sentido, los genéricos cumplen con este último requisito y se convierten en una herramienta para la contención del gasto en medicamentos (37).

De forma general, la estructura del mercado del medicamento de prescripción ambulatoria se representa en la Figura 14, atendiendo a los distintos agentes que intervienen y a la relación existente entre ellos.

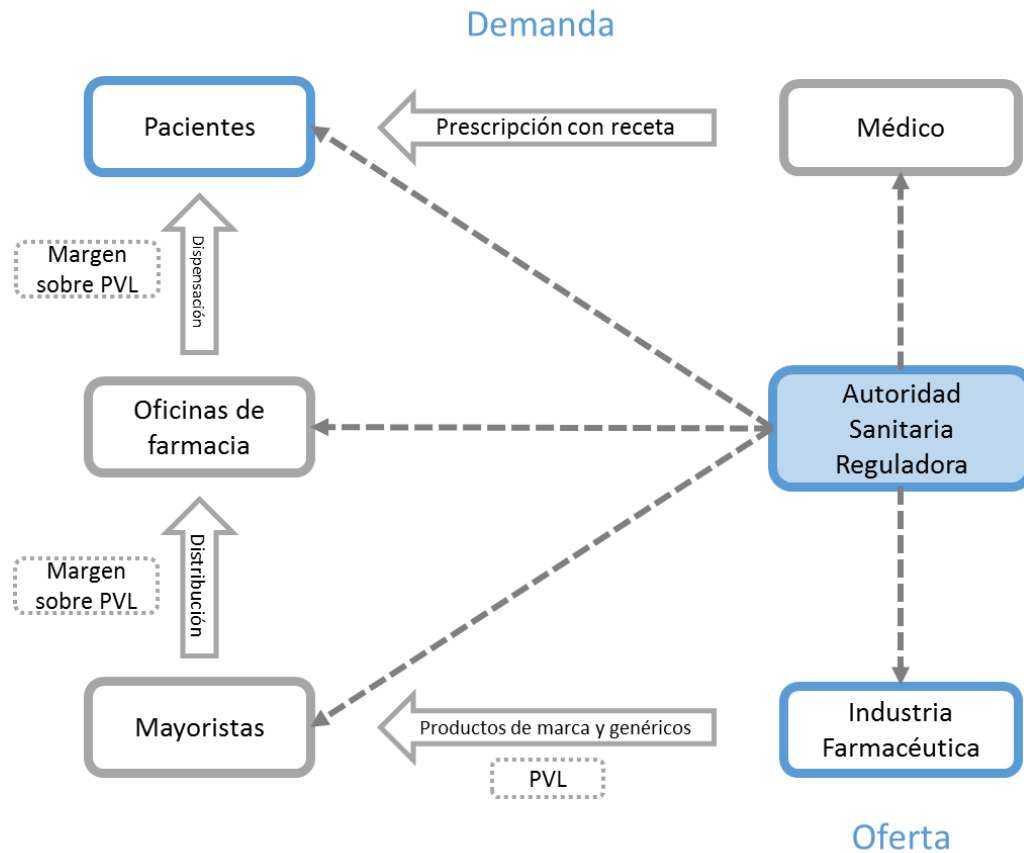


Figura 14. El mercado del medicamento y sus principales agentes

Las características de los componentes del mercado del medicamento de prescripción que acabamos de ver son las siguientes (38):

- **Demanda:** formada por los pacientes o, en este caso, los "consumidores" de medicamentos. Hay que destacar que en este caso es el médico quien hace las veces de "agente" del paciente, y del asegurador, ya que es él quien posee información sobre la efectividad y calidad del medicamento prescrito asociado al problema de salud del paciente, una vez realizado el diagnóstico. Este hecho representa una característica diferencial de este mercado.
- **Oferta:** protagonizada por la industria farmacéutica (en situación de oligopolio).
- **Distribuidores:** los almacenes mayoristas y las oficinas de farmacia (minoristas).
- **Regulador del mercado y del gasto farmacéutico a través de recetas:** la administración central y autonómica
- **Financiador:** hay dos tipos que son el público y el privado

En la Tabla 19 se plantea una posible taxonomía (39) de objetivos perseguidos en la ordenación económica de los medicamentos por parte de la autoridad sanitaria reguladora así como de un conjunto de herramientas posibles en función de los objetivos y organizada en función de los ejes de la oferta y la demanda.

| Eje | Objetivo de la política | Herramientas posibles |
|---------|--|---|
| Oferta | Garantizar la competitividad del mercado | Control de prácticas anticompetitivas |
| | | Promoción de genéricos |
| | Garantizar precios asequibles | Control de precios y márgenes de comercialización |
| | | Negociaciones de precios |
| Demanda | Contención del gasto público en medicamentos | Promoción del uso de genéricos |
| | | Cambios en las modalidades de contratación y adquisición |
| | | Fijación de precios |
| | | Copago |
| | Garantizar la eficiencia del gasto en medicamentos | Evaluación económica de medicamentos y financiación selectiva |

Tabla 19. Alternativas en la regulación económica de los medicamentos

La Tabla 19 también recoge la visión de otros autores sobre los objetivos perseguidos por los mecanismos reguladores (40).

La mayor parte de los países europeos han desarrollado medidas para controlar el gasto farmacéutico y aumentar la eficiencia de su aplicación. Las más significativas manifiestan a continuación (41), (42):

Medidas que afectan a la oferta del mercado farmacéutico más comunes en Europa:

- Listas positivas de medicamentos: son aquellos medicamentos que son parcial o totalmente reembolsables por la Administración.
- Listas negativas de medicamentos: son aquellos medicamentos excluidos de la financiación pública.

- Control de beneficios: bajo esta medida, el Sistema Nacional de Salud determina el margen de los beneficios de la industria farmacéutica derivados de la venta de medicamentos prescritos que está dispuesto a financiar. Si los beneficios de la industria superan dicho límite, estos deberán ser revertidos al Estado.
- Control del precio: la cantidad que financia el Estado depende del precio, de la proporción reembolsable del mismo y del copago. En Europa existen tres tipos de control:
 - El precio de un medicamento debe ser negociado con la Administración antes de ser comercializado (España y Bélgica).
 - El precio se establece de tal manera que se garantice que el rendimiento del capital se mantenga dentro de los límites establecidos (Reino Unido).
 - El precio máximo de un medicamento se determina como la media de los precios del mismo producto en los países vecinos (Grecia, Irlanda, Holanda y Portugal).

Otra forma de control de precios se presenta en forma de acuerdo entre la Administración y la industria sobre el precio y el volumen de venta. En él, el laboratorio farmacéutico establece sus previsiones de venta del medicamento en cuestión de manera que si esta previsión es sobrepasada, el laboratorio es penalizado con una reducción del precio. El objetivo de esta medida es evitar las estrategias de algunos laboratorios que hacían la estimación de un volumen de ventas muy reducido con el objeto de establecer precios más elevados.

Se da la circunstancia (38), que en aquellos países donde se establecen controles de precios, el gasto farmacéutico con respecto al PIB y al gasto sanitario son iguales o superior al resto. En los países con control establecido sobre los precios, el aumento del gasto viene dado por una mayor cantidad consumida, o por la introducción y mayor consumo de productos novedosos, generalmente más caros.

- Precios de referencia: es el máximo precio financiable por la Administración. Se define mediante el establecimiento de precios de referencia para los

medicamentos fuera de patente. Este sistema consiste en agrupar los medicamentos similares en grupos homogéneos en base a su principio activo; y para cada grupo, el financiador determina la cantidad máxima que está dispuesto a financiar. En el plano operativo, el farmacéutico está obligado a sustituir el medicamento prescrito por el médico por un medicamento genérico o un medicamento del grupo homogéneo con precio igual o inferior a éste. Dado que en este caso es necesario el consentimiento del paciente, si el paciente prefiere la especialidad prescrita, entonces deberá abonar la diferencia entre su precio y el precio de referencia, más el copago correspondiente al precio de referencia. Esta medida puede verse tanto desde el lado de la oferta como de la demanda.

- Introducción de medicamentos genéricos: son aquellos medicamentos cuya patente del principio activo ha caducado y tienen el mismo principio activo, forma farmacéutica, composición y bioequivalencia que otra especialidad de referencia. Su precio es menor con respecto a las especialidades de marca.
- Utilización de criterios de coste-efectividad: algunos países han introducido criterios de coste-efectividad para la financiación pública de nuevos medicamentos.
- Medidas establecidas sobre los prescriptores: son medidas que tienen como objetivo trasladar la responsabilidad del gasto a los médicos, incentivando determinados patrones de prescripción. Su base argumental es que la eficiencia en el gasto puede conseguirse a través de la actuación del médico. Las medidas pueden ser de tipo monetario, educativas o de información.

Con respecto a las medidas que afectan a la demanda del mercado farmacéutico en Europa, destaca el establecimiento del copago (parte del precio del medicamento que debe pagar el paciente en el momento de su adquisición en la oficina de farmacia). Su efectividad depende de dos factores (38):

- De la forma en que sea establecido:
 - Tasa fija por receta o por medicamento.
 - Como porcentaje del valor del medicamento prescrito.

- Como una deducción hasta alcanzar cierto límite.
 - Como una combinación de alguna de las anteriores.
- Del número total de prescripciones que están exentas de pago, es evidente, que cuanto mayor sea la población exenta de pago, menor efectividad tiene esta medida. Por ello, suele ir acompañada de medidas complementarias que limiten su gasto.

Respecto a la oferta en el mercado de prescripción, protagonizada por laboratorios farmacéuticos, cabe destacar los siguientes puntos clave (43):

- El aumento del índice de concentración de las empresas en el mercado de prescripción, con especial énfasis en los últimos años.
- Disminución en los últimos años del número de lanzamientos al mercado de nuevas moléculas.
- Las considerables barreras de entrada a la industria farmacéutica de medicamentos de prescripción innovadores. Debido, entre otros, a los siguientes factores:
 - Existencia de patentes.
 - Necesidad de elevadas inversiones en investigación y desarrollo.
 - Elevados gastos de comercialización del producto y marketing.
 - Necesidad de una elevada imagen de compañía como factor de credibilidad, seguridad y confianza para el médico prescriptor.

La industria farmacéutica tiene un papel fundamental en el mercado de medicamentos con prescripción médica por tratarse de un mercado oligopolístico (muchos productos se desarrollan bajo patentes) que les confiere cierto poder de mercado. Asimismo, las empresas farmacéuticas pueden establecer el PVL (Precio de Venta del Laboratorio) para los nuevos medicamentos "de marca" por encima del coste marginal debido a la existencia de altos costes de I+D (38).

La regulación de precios es un fenómeno generalizado en todos los países (44), en un principio, en respuesta a los seguros que mina la sensibilidad de los consumidores a los

precios y crea poderosos pagadores gubernamentales, y también en respuesta a la preocupación de la salud pública en el sentido de que los medicamentos sean asequibles. Debido a que los detalles de estas relaciones regulatorias entre las aseguradoras y los agentes de los pacientes difieren de un país a otro, el papel de la política antimonopolio y de defensa de la competencia también difiere en cierta medida entre los países, a pesar de que los principios fundamentales son similares.

Además, como ya hemos comentado antes, se dan circunstancias como que la relación que existe entre el paciente o asegurador y el médico es imperfecta, que el médico tiene falta de información sobre alternativas más baratas, la existencia de lealtad al producto de marca (sobre la que el médico ha adquirido confianza durante el periodo de patente) y la falta de incentivos para cambiar los hábitos de prescripción, pueden conformar elementos que de manera indirecta otorgan a los laboratorios farmacéuticos la capacidad de establecer precios elevados.

Mercado de medicamentos genéricos

Según la Ley 29/2006, de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios (20), un genérico es "todo medicamento que tenga la misma composición cualitativa y cuantitativa en principios activos y la misma forma farmacéutica, y cuya bioequivalencia con el medicamento de referencia haya sido demostrada por estudios adecuados de biodisponibilidad". La introducción de los medicamentos genéricos en España está en línea con el uso racional del medicamento promulgado por la Organización Mundial de la Salud como ya comentamos anteriormente.

Un medicamento genérico puede comercializarse a partir del fin de la patente del medicamento innovador; esto es, a partir de los diez años de exclusividad por parte del laboratorio descubridor de la fórmula. Pueden ser desarrollados, fabricados, distribuidos y comercializados por cualquier laboratorio farmacéutico que cumpla la normativa legal vigente.

Las principales diferencias con respecto a los medicamentos bajo patente son (37):

- El precio, considerablemente más bajo que el de los medicamentos de marca.
- El público objetivo de la industria farmacéutica (oferta): además de la Administración y los médicos de atención primaria y hospitalaria, el público objetivo se amplía a oficinas de farmacia y pacientes.
- Su utilización como medida de control del gasto farmacéutico y esto lo convierte en un medicamento “diferenciado”.

La cuota de medicamentos genéricos en el mercado farmacéutico total, sea en unidades vendidas o en valores, coloca a España lejos de la media europea que está en torno al 55% en unidades y al 25% en valores (45).

A nivel de prescripción, el 38% de los medicamentos españoles son genéricos, mientras que la cuota europea ronda el 60% (46).

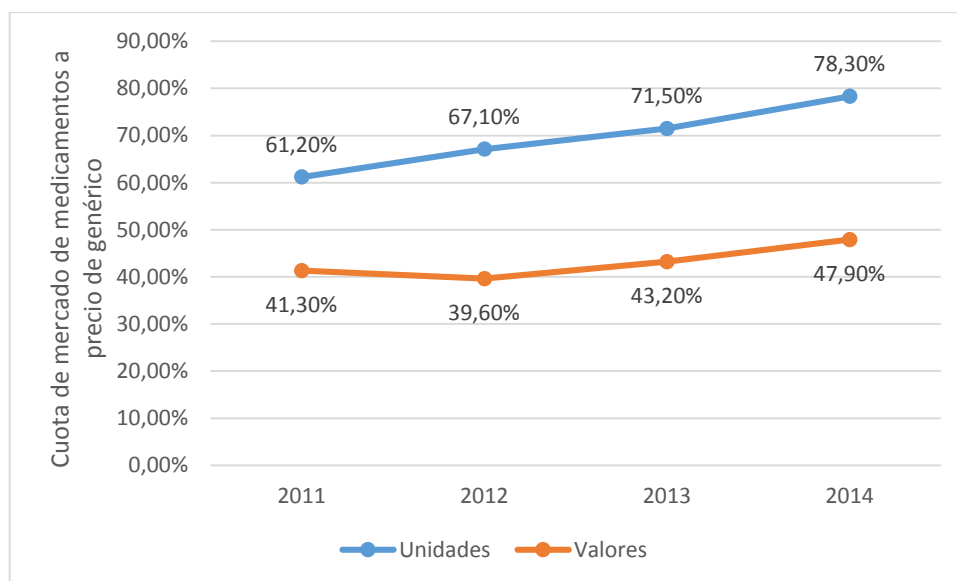


Figura 15. Cuota de mercado de medicamentos a precio de genérico

Fuente: Farmaindustria “Análisis de la evolución del gasto farmacéutico público en España”. Enero 2015 (47)

Desde un punto de vista de la dispensación, podemos observar en la Figura 15 que si tenemos en cuenta las unidades de medicamentos dispensadas en el año 2014, el

78,30% son medicamentos genéricos mientras que este porcentaje baja al 47,90% si consideramos la cuota sobre el valor de las ventas en euros.

Según *Health Market Research* (hmR) (48), datos aparecen en la Tabla 20, el año 2014 presentó cifras positivas respecto al mercado farmacéutico total. Según dicho informe, este mercado experimentó un crecimiento del 0,3% respecto a 2013, generando 17.600 millones de euros y registrando, además, un crecimiento del 1,5% en unidades. Los autores del estudio destacan el hecho de que estas cifras positivas se obtienen pese a la caída del segmento de medicamentos de marca, que lo hace un 1,6% en valores y un 2,8% en unidades.

| Autonomía | Comparación 2014 con respecto a 2013 | | | | | |
|-----------------------|--------------------------------------|---------------|--------------|-----------------------|---------------|--------------|
| | VALORES | | | UNIDADES | | |
| | Total prescripción | Marca | Genérico | Total prescripción | Marca | Genérico |
| Andalucía | -5,20% | -9,40% | 9,50% | -2,90% | -10,50% | 6,10% |
| Aragón | -0,10% | -0,60% | 2,30% | 1,70% | -0,70% | 5,20% |
| Asturias | 3,50% | 1,70% | 13,20% | 4,00% | 0,70% | 10,10% |
| Baleares | 0,70% | -1,30% | 9,40% | 2,90% | -1,10% | 7,90% |
| Canarias | 4,60% | 4,90% | 3,50% | 2,20% | 0,40% | 4,90% |
| Cantabria | 3,60% | 3,60% | 3,90% | 3,10% | 2,10% | 4,80% |
| Castilla la Mancha | -0,80% | -1,80% | 2,70% | 0,00% | -2,80% | 3,50% |
| Castilla y León | 4,00% | 2,00% | 10,50% | 3,80% | -30,00% | 7,90% |
| Cataluña | 1,50% | -0,10% | 7,70% | 2,50% | -1,70% | 7,30% |
| Comunidad Valenciana | -0,60% | -1,50% | 3,80% | 1,50% | -0,70% | 5,70% |
| Extremadura | 4,60% | 3,50% | 11,00% | 4,00% | 1,30% | 9,20% |
| Galicia | 0,50% | -1,20% | 8,70% | 2,20% | -2,50% | 10,50% |
| La Rioja | 4,50% | 4,10% | 6,70% | 3,70% | 1,00% | 8,40% |
| Madrid | 0,80% | -1,50% | 9,30% | 1,20% | -3,50% | 7,00% |
| Murcia | -6,90% | 6,30% | 25,40% | 12,20% | 6,80% | 23,90% |
| Navarra | 0,50% | -8,20% | -1,20% | -5,40% | -8,00% | -1,30% |
| País Vasco | 0,30% | -0,80% | 4,90% | 1,60% | 0,00% | 3,30% |
| Total nacional | 0,30% | -1,60% | 7,80% | 1,30% | -2,80% | 6,90% |

Tabla 20. Comparación del segmento de genéricos del año 2014 respecto del 2013 en unidades y valores por comunidades autónomas

Fuente: HEALTH MARKET RESEARCH ESPAÑA

Si se analiza el comportamiento de los medicamentos genéricos por comunidades autónomas, Tabla 20, se observa que el único descenso se presenta en Navarra tanto en

unidades como en valores. Las subidas más destacadas las presentan Murcia (25,40%), Asturias (13,20%), Extremadura (11,00%) y Castilla y León (10,50%).

Si se analiza la situación de los medicamentos de marca, se observa que las mayores caídas se produjeron en Andalucía (-9,4%), Navarra (-8,25%) y Castilla La Mancha (-1,8%). En el otro extremo, este mismo segmento consigue sus mejores resultados en Murcia (+6,35%), Canarias (+4,9%) y La Rioja (+4,1%).

La integración de ambos análisis nos muestra que el segmento de los genéricos ha amortiguado el descenso del segmento de los medicamentos de marca con respecto al comportamiento general del mercado de medicamentos de prescripción.

| Tasa de variación interanual (diciembre 2014 respecto diciembre 2013) | | |
|--|----------------|-----------------|
| | Valores | Unidades |
| Total mercado | 3,10% | 3,20% |
| Total prescripción | 3,90% | 5,00% |
| Marca | 3,10% | 1,80% |
| Genéricos | 6,90% | 9,10% |
| Total Consumer Health | 1,00% | -0,90% |

Tabla 21. Tasa de variación interanual (diciembre 2014 respecto diciembre 2013)

Fuente: HEALTH MARKET RESEARCH ESPAÑA

Si se analiza solo diciembre de 2014 respecto al mismo mes del año anterior, Tabla 21, el mercado farmacéutico creció un 3,1%, con cifras positivas en todos los segmentos: marca (3,1%), genéricos (6,9%) y *Consumer Health* (1,0%). De forma gráfica podemos observarlo en la Figura 16.

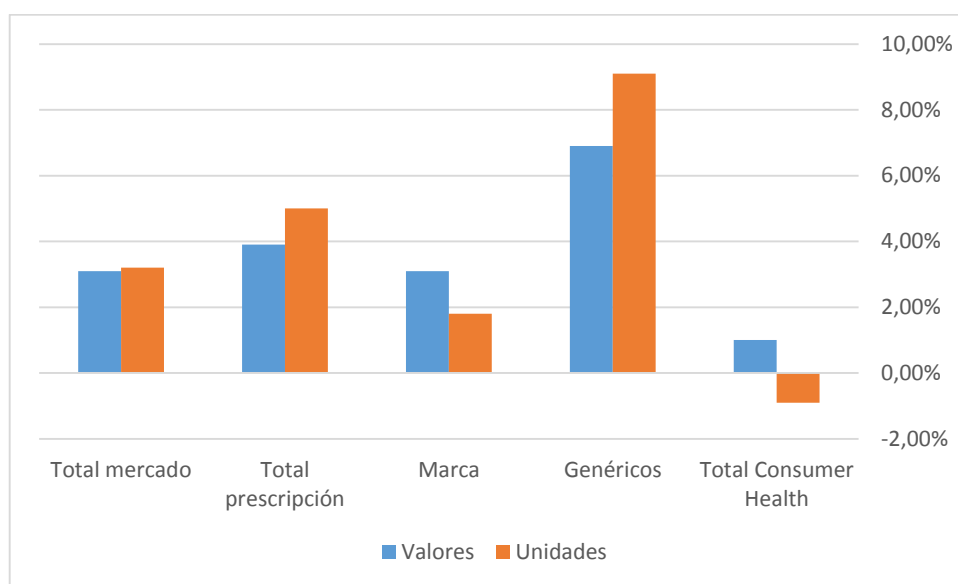


Figura 16. Tasa de variación interanual (diciembre 2014 respecto diciembre 2013)
Fuente: HEALTH MARKET RESEARCH ESPAÑA

Con independencia de los datos más recientes que hemos analizado, la tendencia del mercado muestra que al contrario que ocurre con los medicamentos innovadores o de marca, el valor del segmento de genéricos ha crecido un 74% en los últimos cinco años a tenor de los datos ofrecidos por el último informe de la consultora *IMS Health* (47). En dicho estudio se asegura que en el total del mercado farmacéutico la cuota de mercado de los genéricos alcanzó en marzo del año 2015 el 39% en unidades.

Como podemos ver a lo largo de este trabajo, existen multitud de fuentes de información sobre un mismo hecho y debemos tener en cuenta el máximo posible de ellas para disminuir la incertidumbre sobre el aspecto analizado.

En la Figura 17 se refleja el impacto que produce una fuerte implantación de los medicamentos genéricos sobre el ahorro sanitario. Así, en el Reino Unido, el ahorro anual en el año 2010 generado por los medicamentos genéricos ascendió a 6.000 millones de euros, según los últimos datos de la Asociación Europea de Medicamentos Genéricos (EGA) (45). Este hecho avala la puesta en marcha de la prescripción de genéricos por parte de los países más desarrollados.

En el caso de España cabría analizar si el menor ahorro se debe al menor precio de los medicamentos en general, en comparación con otros países europeos.

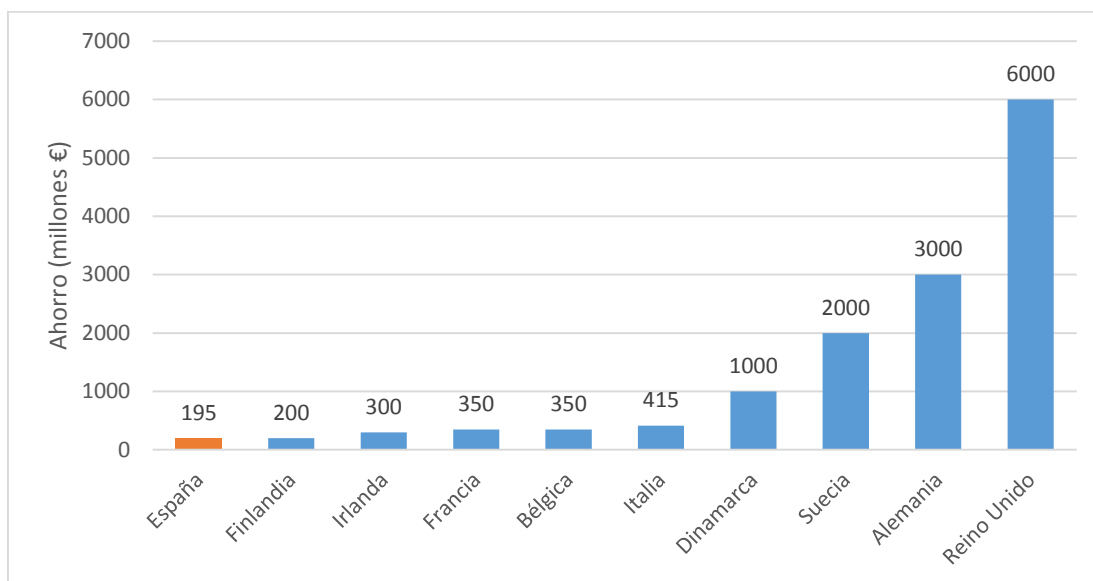


Figura 17. Ahorro provocado por el uso de genéricos (millones de €). Año 2010

Medicamentos hospitalarios

Estos medicamentos son prescritos en los hospitales del sistema público. Su dispensación es realizada por los servicios de farmacia hospitalaria.

Por lo general, los medicamentos que se prescriben, se preparan y se administran en un hospital son los mismos que se prescriben en el ámbito ambulatorio de los centros de salud de atención primaria. Aunque existen ciertos medicamentos de uso exclusivo hospitalario.

La alta dotación tecnológica de los hospitales y su orientación especialista, hace que en él se desarrollen actividades de alto interés para los laboratorios farmacéuticos (49):

1. Ensayos de nuevos productos y protocolos de uso de medicamentos.
2. Es un foco de irradiación de conocimiento a través de protocolos y pautas que son seguidas por el colectivo médico tanto dentro como fuera del hospital.
3. Se evalúan medicamentos por parte de líderes de opinión que ensayan y publican el beneficio terapéutico del medicamento.

En la Figura 18 (49), se puede observar los proveedores que tiene el hospital para el abastecimiento de medicamentos y productos médicos de diagnóstico.

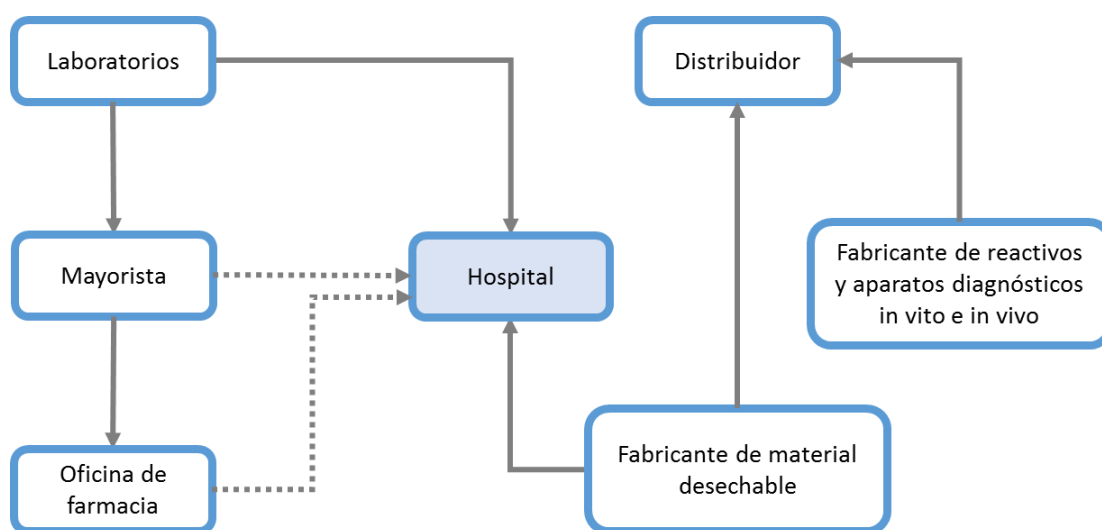


Figura 18. Acceso de medicamentos, productos médicos y de diagnóstico al hospital

En el apartado sobre el gasto farmacéutico público que veremos más adelante, se analiza la evolución de este mercado.

1.2.5.1.1 Prestación farmacéutica pública

En España, la prestación farmacéutica comprende los medicamentos y productos sanitarios y el conjunto de actuaciones encaminadas a que los pacientes los reciban de forma adecuada a sus necesidades clínicas, en las dosis precisas según sus requerimientos individuales, durante el período de tiempo adecuado y al menor coste posible para ellos y para la comunidad (50).

Esta prestación se rige por lo dispuesto en la Ley 29/2006, de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios y demás disposiciones aplicables.

La cartera de servicios comunes de prestación farmacéutica está compuesta por (50):

1. En el caso de pacientes no hospitalizados, la prestación farmacéutica:
 - Comprende la indicación, prescripción y dispensación de los siguientes productos:

- Los medicamentos para los que, de acuerdo con la normativa vigente, se resuelva su financiación y condiciones de dispensación en el Sistema Nacional de Salud y que hayan sido autorizados y registrados por la Agencia Española de Medicamentos y Productos Sanitarios o de acuerdo con lo dispuesto por las normas europeas que establecen los procedimientos comunitarios para la autorización y control de los medicamentos de uso humano.
 - Los efectos y accesorios que dispongan del correspondiente marcado CE y para los que, de acuerdo con la normativa vigente, se resuelva su financiación y condiciones de dispensación en el Sistema Nacional de Salud.
 - Las fórmulas magistrales y los preparados oficinales elaborados por las oficinas de farmacia de acuerdo con lo establecido en el Formulario Nacional y que cumplan las normas de la Real Farmacopea Española, así como la normativa vigente sobre las normas para su correcta elaboración y control de calidad, en las condiciones pactadas en los correspondientes conciertos suscritos entre las administraciones sanitarias autonómicas y las oficinas de farmacia.
 - Las vacunas individualizadas antialérgicas y las vacunas individualizadas bacterianas, preparadas con agentes inmunizantes, a concentración y dilución específica en base a la correspondiente prescripción facultativa para un paciente determinado, de acuerdo con la normativa vigente.
- Excluye:
 - Los productos de utilización cosmética, dietéticos, aguas minerales, elixires, dentífricos y otros productos similares.
 - Los medicamentos calificados como publicitarios.
 - Los medicamentos adscritos a los grupos o subgrupos terapéuticos excluidos de la financiación por la normativa vigente.

- Los medicamentos homeopáticos.
 - Los efectos y accesorios de los que se realice publicidad dirigida al público en general.
2. En el caso de pacientes hospitalizados, la prestación farmacéutica comprende los productos farmacéuticos que necesiten los pacientes conforme se recoge en la cartera de servicios comunes de atención especializada (51).

Dichos productos se prescriben y dispensan en la asistencia ambulatoria, facturándose al Sistema Nacional de Salud a través de farmacias, o bien se prescriben y se administran en hospitales considerándose un gasto propio de los mismos.

A continuación, tratamos la evolución de estos mercados desde el punto de vista económico.

1.2.5.1.2 Gasto farmacéutico público

El objetivo de una política de medicamentos es asegurar que están disponibles medicamentos seguros, eficaces y de calidad, para cubrir las necesidades sanitarias de un país.

El gasto que se produce en España por la prestación farmacéutica del Sistema Nacional de Salud, conocido como gasto farmacéutico, presenta una serie de características (52):

- Es un componente importante del gasto sanitario.
- Tradicionalmente incontrolado, está más controlado en los últimos años.
- Las medidas para reducirlo son similares a las de otros países de Europa.
- Presenta un amplio rango de variación entre las CCAA.
- La tendencia de crecimiento ahora es convergente con el gasto sanitario no farmacéutico público.
- Es superior a la media de la Unión Europea.
- La información no es transparente.
- No se consumen suficientes genéricos.

- La I+D de la Industria farmacéutica se puede considerar insuficiente.
- Existe un lobby poderoso que trata de condicionar las decisiones de administraciones y políticos.

El gasto farmacéutico público presenta dos componentes: el gasto farmacéutico público en recetas y el gasto farmacéutico público hospitalario.

El gasto farmacéutico público en recetas en España se corresponde con la facturación de recetas médicas del Sistema Nacional de Salud (CCAA e INGESA: Instituto Nacional de Gestión Sanitaria). Este componente del gasto farmacéutico tiene un mayor impacto sobre las oficinas de farmacia por lo que se analizará con más detalle en esta investigación.

En la información disponible de gasto farmacéutico, no se ha considerado el copago realizado por los usuarios en las prescripciones ambulatorias (52).

Gasto farmacéutico público hospitalario

En el caso hospitalario, es de interés comentar como explicación a la diversidad de fuentes de información, que el gasto farmacéutico público hospitalario se corresponde con la operación estadística denominada “Estadística de Gasto Farmacéutico de los Hospitales Públicos del Sistema Nacional de Salud” (53) que está incluida en el inventario de operaciones estadísticas del Instituto Nacional de Estadística desde el año 2010, y que lamentablemente aún no dispone de datos por lo que como fuente estadística pública oficial no puede ser utilizada en esta investigación.

En cualquier caso, y de cara a poder realizar el proceso de toma de decisiones fiable por todos los agentes del mercado como garante de una correcta gestión, la transparencia entendida como disponibilidad y accesibilidad de los datos, se torna fundamental (54).

Como alternativas hay dos fuentes reconocidas en el sector que son las siguientes:

- *IMS Health*: sus datos provienen de un panel de 139 hospitales. Vienen expresados a precios PVL, aunque no es exactamente el precio al que compran los hospitales (55).

- Farmaindustria: sus datos provienen de la Encuesta de Deuda Hospitalaria realizada por esta asociación (56).

Ambas fuentes presentan algunas diferencias entre ellas en lo que respecta a la estimación del gasto farmacéutico hospitalario, así que utilizaremos ambas en esta investigación y veremos si hay homogeneidad en sus estimaciones antes de su interpretación.

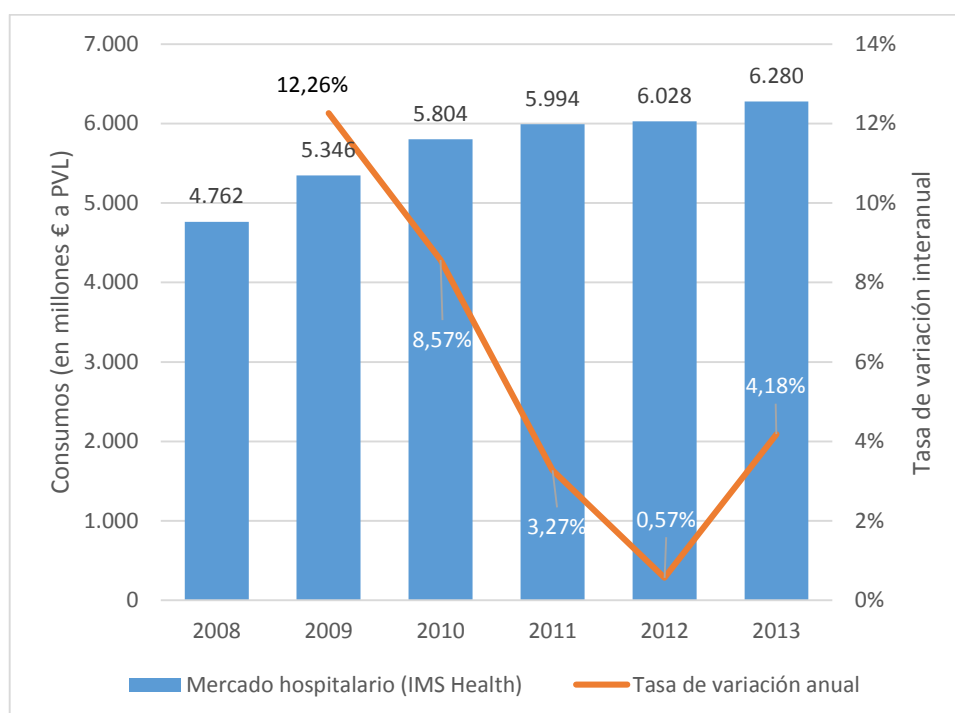


Figura 19. Mercado hospitalario español según IMS Health

Fuente: IMS Health "Análisis del mercado hospitalario 2008- 2012" y actualización 2013. Precios PVL

Como se aprecia en la Figura 19, el crecimiento del mercado hospitalario español (según *IMS Health*) aunque se mantiene positivo, su incremento es año tras año menor que el anterior en el período considerado 2008-2012. Aun así, comienza acelerar su crecimiento en 2013 después de tres años de ralentización.

Si utilizamos los datos de Farmaindustria (Figura 20), las conclusiones son parecidas en lo que respecta a la consideración de la variación anual. Si bien, según esta fuente, en términos absolutos el mercado es menor del que plantea *IMS Health*.

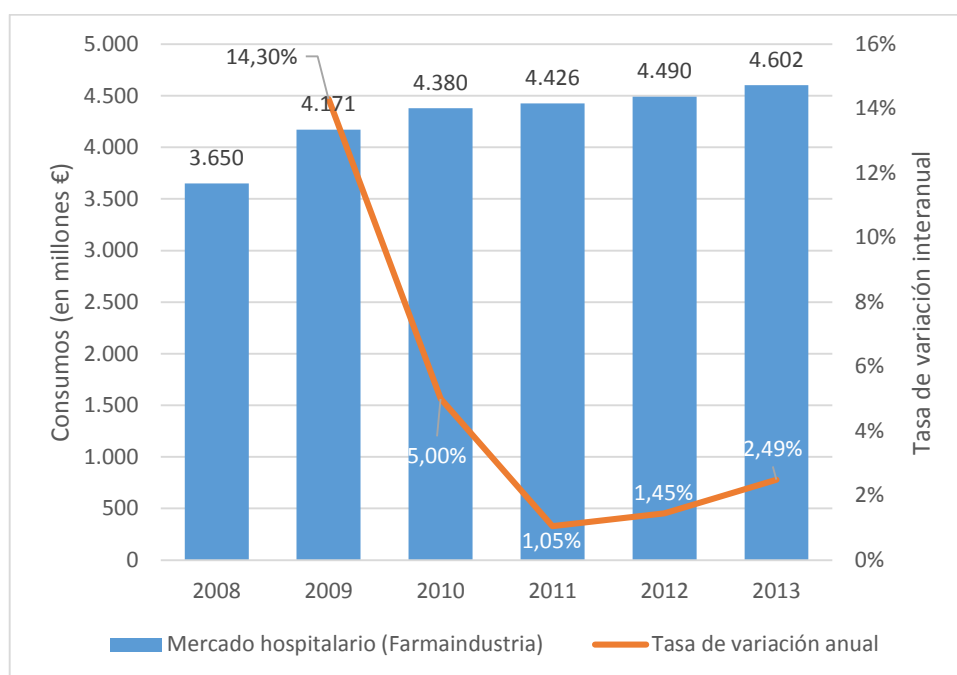


Figura 20. Mercado hospitalario español según Farmaindustria

Fuente: Farmaindustria "Set Indicadores Gasto Farmacéutico Público" a partir de la Encuesta de Deuda Hospitalaria. 2015

Ambas fuentes, con distribución similar, coinciden en que la evolución del mercado en el periodo 2008-2013 presenta unas tasas de variación interanual muy suaves, lo que confiere al mercado un crecimiento muy moderado. Caso distinto a lo que veremos más adelante en la evolución del gasto farmacéutico público en recetas del Sistema Nacional de Salud.

No obstante, al no ser transparente no se pueden analizar los factores que influyen en el mismo (57).

Gasto farmacéutico público en recetas del Sistema Nacional de Salud

La factura en medicamentos de la Administración Pública (Estado y CCAA) en España, a través de las oficinas de farmacia, viene dada por la cantidad y el tipo de medicamentos prescritos, bien de marca o genéricos, más la parte del precio que financia la propia administración. Sobre estas variables, las administraciones de la mayor parte de los países occidentales han actuado, bien reduciendo o controlando el precio (oferta) o bien tratando de controlar la cantidad y la tipología de los medicamentos prescritos (demanda) (39), como pudimos observar en la Tabla 19.

Es opinión del sector, bajo el prisma de que el gasto en sí mismo no es malo sino que es el gasto ineficiente el que debería ser evitado, que la identificación de posibles mejoras se consigue a través de opiniones expertas. En este sentido, es necesario considerar la efectividad de los medicamentos prescritos, el efecto en el bienestar y la salud de los ciudadanos y, por último, en qué medida el gasto en medicamentos puede ser reducido sin que dicho bienestar se vea alterado (38).

El gasto farmacéutico público en recetas es de las pocas magnitudes estudiadas dentro del mercado farmacéutico español que tiene una fuente pública bien determinada y con una trayectoria de varios años que permite disponer de una serie temporal y por tanto analizar su tendencia.

Las variables que se utilizan en la medición del gasto farmacéutico público en recetas son las siguientes:

- Recetas facturadas del Sistema Nacional de Salud (R): es el número de recetas dispensadas en las oficinas de farmacia cuya prescripción ha sido realizada por un médico del Sistema Sanitario Público.
- Gasto farmacéutico a través de receta oficial del Sistema Nacional de Salud (G): en la suma de los importes de las recetas dispensadas en las oficinas de farmacia cuya prescripción ha sido realizada por un médico del Sistema Sanitario Público. No contabiliza la aportación que hace el usuario según sea su caso.

- Gasto medio por receta facturada del Sistema Nacional de Salud: se trata del cociente entre los dos indicadores anteriores: $\text{Gasto medio por receta} = G/R$.

Para cada uno de estos indicadores hemos calculado una serie de medidas estadísticas para la comparación de los mismos. Dichas medidas son:

- La tasa de variación media acumulada (TVMA).
- Media aritmética anual de la serie para cada Comunidad.
- Cuota media anual de la serie para cada Comunidad sobre el total nacional.
- Cuota del año 2014 para cada Comunidad sobre el total nacional.

Los datos han sido tomados de la web del Ministerio de Sanidad, Servicios Sociales e Igualdad (56). Hemos optado por la serie 2008-2014 con objeto de normalizar el periodo de referencia de todas las fuentes consultadas. En el caso del gasto farmacéutico público hay datos publicados desde 2003.

A nivel nacional, podemos observar en la Tabla 22 como el número de recetas tuvo hasta el 2011 un comportamiento creciente y a partir de dicho año comienza el decrecimiento. Gracias a la TVMA observamos que en el periodo 2008-2014 decrece un 0,40% anualmente de forma acumulada. Si nos apoyamos en la Figura 21 para un mejor análisis, vemos que se producen dos caídas muy bruscas en los años 2012 y 2013, y una pequeña subida del 2014 que suavizan la TVMA del periodo analizado. Atendiendo al lado de la “demanda” observamos que la prescripción médica se ha visto afectada por la implantación de medidas de control del gasto farmacéutico público que ya se comentó en apartados anteriores. Concretamente, los informes sectoriales (57) relacionan estas caídas con la implantación de los decretos RD 09/2011 (58) y RD 16/2012 (61).

| | Número de Recetas | Gasto farmacéutico por Recetas | Gasto medio por Receta | Población | Gasto Farmacéutico en Recetas por habitante |
|--|----------------------|--------------------------------------|---------------------------------|------------|---|
| 2008 | 890.040.069 | 11.970.955.447 | 13,45 | 46.157.822 | 259,35 |
| 2009 | 934.002.083 | 12.505.692.910 | 13,39 | 46.745.807 | 267,53 |
| 2010 | 957.694.628 | 12.207.683.146 | 12,75 | 47.021.031 | 259,62 |
| 2011 | 973.211.911 | 11.135.401.976 | 11,44 | 47.190.493 | 235,97 |
| 2012 | 913.813.388 | 9.770.933.367 | 10,69 | 47.265.321 | 206,73 |
| 2013 | 859.576.567 | 9.183.249.747 | 10,68 | 47.129.783 | 194,85 |
| 2014 | 868.630.721 | 9.359.975.385 | 10,78 | 46.771.341 | 200,12 |
| TVMA (2008-2014) | -0,40% | -4,02% | -3,63% | 0,22% | -4,23% |
| TV interanual 2014-2013 | 1,05% | 1,92% | 0,86% | -0,76% | 2,71% |

Tabla 22. Indicadores del gasto farmacéutico en el periodo 2008-2014 a nivel nacional

Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad, Instituto Nacional de Estadística y elaboración propia

Con respecto al indicador “gasto farmacéutico por recetas” se observa un descenso medio anual acumulado del 4,02% en el periodo 2008-2014. En este caso podemos ver que es el gasto quien ha soportado las medidas implantadas. Luego si el número de recetas se ha mantenido en el periodo, apenas un descenso del 0,40%, y ha bajado el gasto, se deduce que lo que ha bajado el precio de los medicamentos entre 2008 y 2014 (Figura 21).

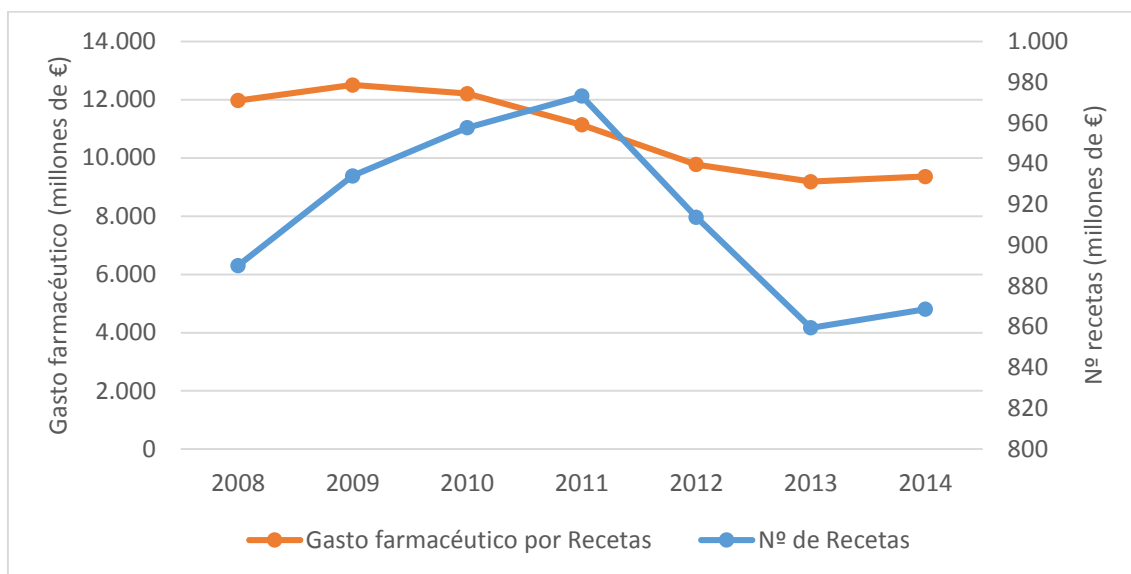


Figura 21. Evolución del número de recetas y gasto farmacéutico por recetas en el periodo 2008-2014

El gasto farmacéutico en recetas por habitante ha evolucionado de la misma manera que el gasto ya que la población se ha mantenido en el periodo estudiado. En el año 2014 se establece en 200,12 € por habitante; lo que ha supuesto una caída de más de 59 € en seis años. El comportamiento de la serie no es tan brusco como el del número de recetas según se observa en la Figura 22.

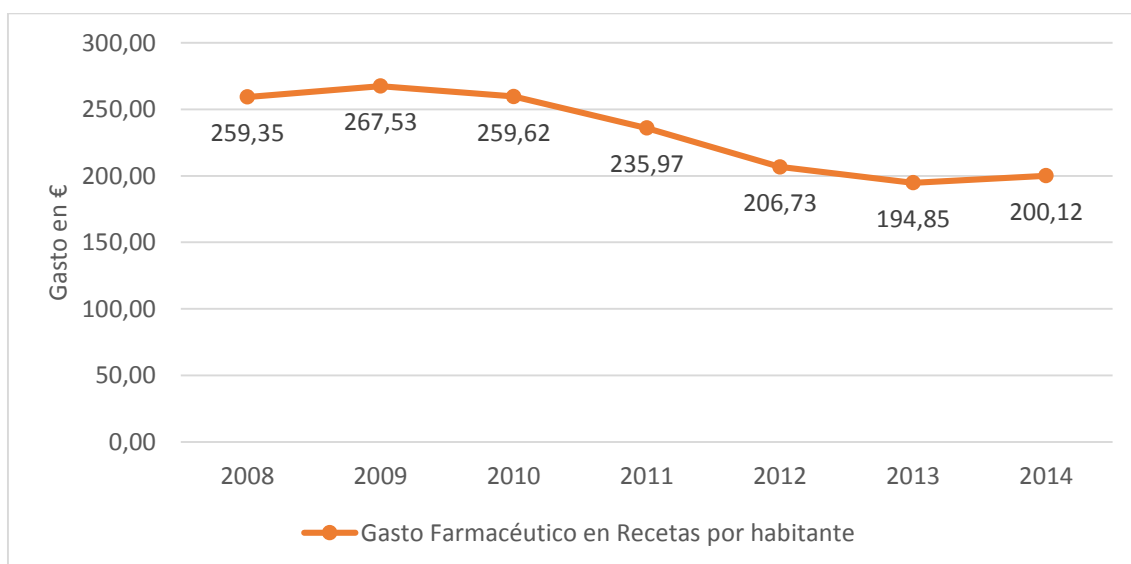


Figura 22. Evolución del gasto farmacéutico por habitante en el periodo 2008-2014

En el caso del gasto medio por receta, se observa un descenso acumulado anual del 3,63% con un comportamiento semejante al gasto farmacéutico por recetas que experimentó un descenso acumulado anual del 4,02% (Figura 23).

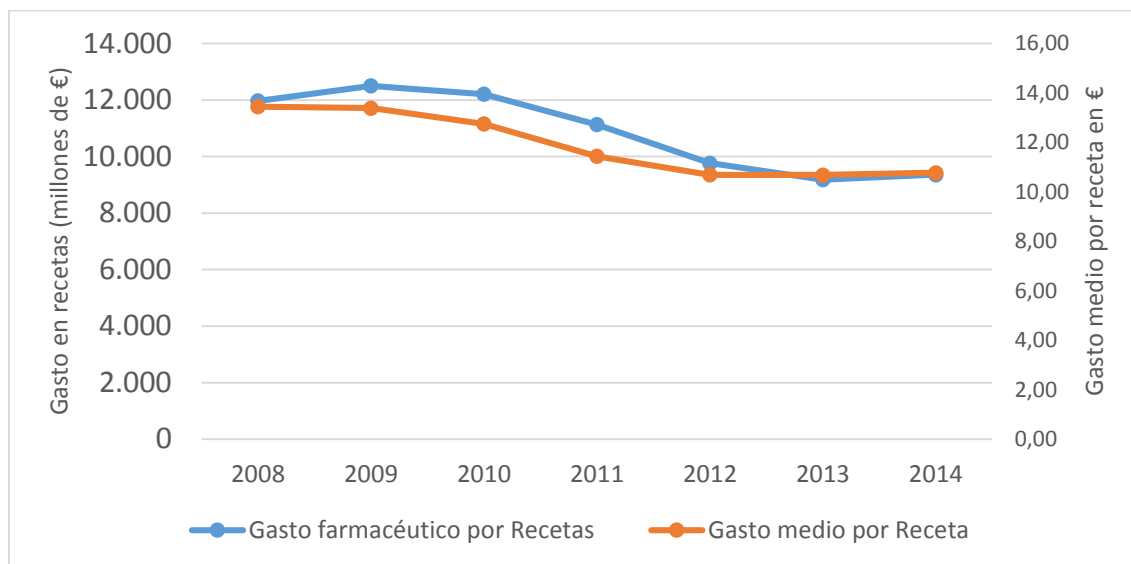


Figura 23. Evolución del gasto farmacéutico por recetas y el gasto medio por receta en el periodo 2008-2014

Si atendemos al 2014 observamos que se produce un aumento en todos los indicadores. Cabe destacar una subida del 1,92% en el caso del gasto farmacéutico por recetas.

Según *IMS Health* (47) las ventas de medicamentos genéricos en el 2014 crecieron un 9,5%, y cayeron un 1,4% las ventas del resto de medicamentos (no Especialidades Farmacéuticas Genéricas).

Dado que las competencias en materia de sanidad están transferidas a las comunidades autónomas, en las tablas que siguen a continuación se desglosa por comunidad el comportamiento de los indicadores estudiados a nivel nacional.

| COMUNIDAD | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | TVMA (2008-2014) | Nº medio de recetas (2008-2014) | Cuota media (2008-2014) | Cuota 2014 |
|-----------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|---------------------|---------------------------------------|----------------------------|----------------|
| Andalucía | 163.230.548 | 173.563.234 | 176.739.209 | 178.913.616 | 170.159.996 | 161.433.846 | 164.251.828 | 0,10% | 169.756.040 | 18,58% | 18,91% |
| Aragón | 27.197.124 | 28.448.972 | 28.940.147 | 29.413.976 | 27.896.280 | 26.019.337 | 26.630.722 | -0,35% | 27.792.365 | 3,04% | 3,07% |
| Asturias | 23.785.167 | 24.514.804 | 24.835.373 | 25.305.044 | 24.051.435 | 22.610.730 | 22.153.216 | -1,18% | 23.893.681 | 2,61% | 2,55% |
| Baleares | 15.299.360 | 15.538.912 | 16.729.643 | 18.112.842 | 17.292.030 | 16.273.864 | 16.743.922 | 1,52% | 16.570.082 | 1,81% | 1,93% |
| Cantabria | 11.196.284 | 11.623.622 | 11.870.895 | 11.644.823 | 11.009.627 | 10.435.569 | 10.613.212 | -0,89% | 11.199.147 | 1,23% | 1,22% |
| Castilla la Mancha | 42.049.940 | 44.376.775 | 45.941.215 | 46.047.461 | 44.283.796 | 41.465.318 | 41.834.696 | -0,09% | 43.714.172 | 4,78% | 4,82% |
| Castilla y León | 49.406.054 | 51.246.512 | 52.744.305 | 53.810.280 | 50.333.121 | 48.043.912 | 48.874.322 | -0,18% | 50.636.929 | 5,54% | 5,63% |
| Cataluña | 143.056.512 | 147.582.927 | 151.486.758 | 154.434.341 | 141.670.319 | 131.798.335 | 134.185.746 | -1,06% | 143.459.277 | 15,70% | 15,45% |
| Canarias | 38.243.802 | 40.819.665 | 41.311.624 | 41.705.661 | 40.089.910 | 38.547.026 | 39.898.478 | 0,71% | 40.088.024 | 4,39% | 4,59% |
| Extremadura | 23.694.465 | 25.608.038 | 26.477.061 | 27.076.118 | 25.917.615 | 24.836.427 | 25.597.945 | 1,30% | 25.601.096 | 2,80% | 2,95% |
| Galicia | 60.117.987 | 63.369.283 | 65.572.028 | 66.677.377 | 60.862.124 | 59.999.811 | 59.960.938 | -0,04% | 62.365.650 | 6,82% | 6,90% |
| Madrid | 97.985.348 | 103.897.368 | 106.826.195 | 110.012.143 | 104.754.914 | 97.066.688 | 98.947.729 | 0,16% | 102.784.341 | 11,25% | 11,39% |
| Murcia | 28.247.808 | 29.742.500 | 30.866.297 | 31.426.208 | 29.640.723 | 27.711.991 | 28.103.426 | -0,09% | 29.391.279 | 3,22% | 3,24% |
| Navarra | 11.341.909 | 11.803.180 | 12.042.936 | 12.297.949 | 11.694.928 | 10.888.274 | 11.269.587 | -0,11% | 11.619.823 | 1,27% | 1,30% |
| Comunidad Valenciana | 106.969.994 | 111.787.286 | 114.020.266 | 114.140.645 | 103.776.378 | 95.557.374 | 96.120.100 | -1,77% | 106.053.149 | 11,61% | 11,07% |
| País Vasco | 40.353.655 | 41.818.572 | 42.872.249 | 43.551.367 | 42.180.035 | 39.152.385 | 35.602.341 | -2,07% | 40.790.086 | 4,46% | 4,10% |
| La Rioja | 5.825.493 | 6.129.407 | 6.212.762 | 6.365.271 | 6.064.446 | 5.712.411 | 5.813.481 | -0,03% | 6.017.610 | 0,66% | 0,67% |
| Ceuta | 1.103.811 | 1.145.372 | 1.183.346 | 1.227.124 | 1.164.177 | 1.091.146 | 1.110.361 | 0,10% | 1.146.477 | 0,13% | 0,13% |
| Melilla | 934.808 | 985.654 | 1.022.319 | 1.049.665 | 971.534 | 932.123 | 918.671 | -0,29% | 973.539 | 0,11% | 0,11% |
| NACIONAL | 890.040.069 | 934.002.083 | 957.694.628 | 973.211.911 | 913.813.388 | 859.576.567 | 868.630.721 | -0,40% | 913.852.767 | 100,00% | 100,00% |

Tabla 23. Número de recetas facturadas del Sistema Nacional de Salud en el periodo 2008-2014. Tasas de variación y reparto según las CCAA

Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad y elaboración propia

En lo que respecta al comportamiento global de la serie, observamos en la Tabla 23 como el País Vasco es la Comunidad que menor TVMA presenta, con un valor de - 2,07%, es decir, que en el periodo 2008-2014 ha experimentado un decrecimiento anual medio acumulado del 2,07% del número de recetas facturadas. Cerca del País Vasco aparece la Comunidad Valenciana, con una TVMA del -1,77%. En el otro extremo aparecen las comunidades de Baleares (+1,52%) y Extremadura (+1,30%).

En la Figura 24 se observa con más claridad el comportamiento del número de recetas facturadas al Sistema Nacional de Salud. En ella observamos como cinco comunidades están por debajo de la media nacional y seis que han aumentado el número de recetas facturadas presentando tasas positivas.

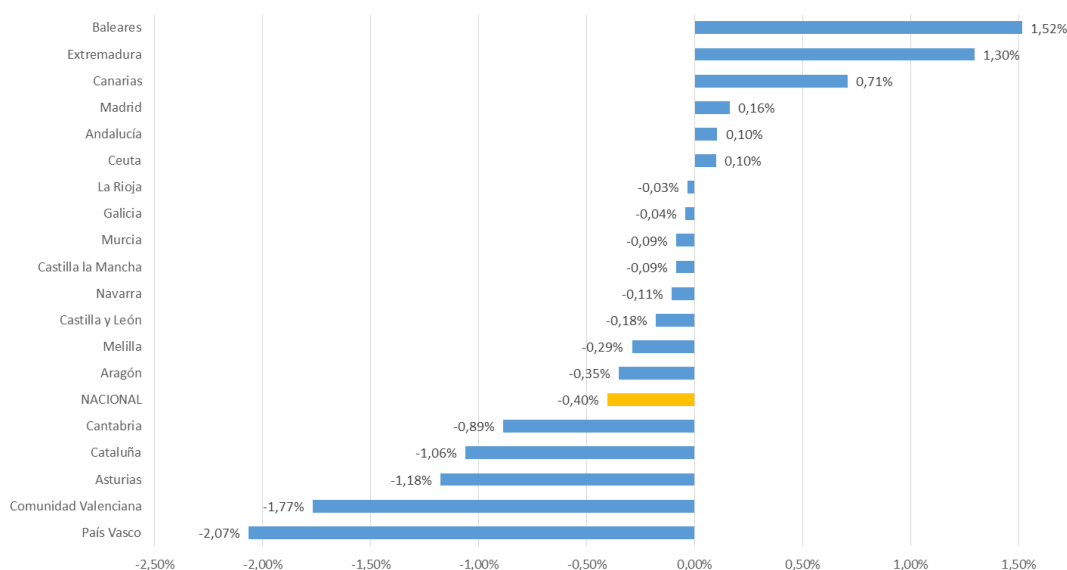


Figura 24. TVMA del número de recetas facturadas del Sistema Nacional de Salud en el periodo 2008-2014

La evolución con respecto al año 2008 la mediremos a través de los números índices simples con base 2008 que aparecen en la Tabla 24 y en la Figura 25. Destaca como a partir del 2011 se produce una divergencia en el comportamiento del número de recetas facturadas. Esto es debido a las diferentes políticas realizadas en materia de contención del gasto público llevadas a cabo por las diferentes comunidades.

| COMUNIDAD | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|----------------------|---------------|---------------|---------------|---------------|---------------|--------------|--------------|
| Andalucía | 100,00 | 106,33 | 108,28 | 109,61 | 104,25 | 98,90 | 100,63 |
| Aragón | 100,00 | 104,60 | 106,41 | 108,15 | 102,57 | 95,67 | 97,92 |
| Asturias | 100,00 | 103,07 | 104,42 | 106,39 | 101,12 | 95,06 | 93,14 |
| Baleares | 100,00 | 101,57 | 109,35 | 118,39 | 113,02 | 106,37 | 109,44 |
| Cantabria | 100,00 | 103,82 | 106,03 | 104,01 | 98,33 | 93,21 | 94,79 |
| Castilla la Mancha | 100,00 | 105,53 | 109,25 | 109,51 | 105,31 | 98,61 | 99,49 |
| Castilla y León | 100,00 | 103,73 | 106,76 | 108,91 | 101,88 | 97,24 | 98,92 |
| Cataluña | 100,00 | 103,16 | 105,89 | 107,95 | 99,03 | 92,13 | 93,80 |
| Canarias | 100,00 | 106,74 | 108,02 | 109,05 | 104,83 | 100,79 | 104,33 |
| Extremadura | 100,00 | 108,08 | 111,74 | 114,27 | 109,38 | 104,82 | 108,03 |
| Galicia | 100,00 | 105,41 | 109,07 | 110,91 | 101,24 | 99,80 | 99,74 |
| Madrid | 100,00 | 106,03 | 109,02 | 112,27 | 106,91 | 99,06 | 100,98 |
| Murcia | 100,00 | 105,29 | 109,27 | 111,25 | 104,93 | 98,10 | 99,49 |
| Navarra | 100,00 | 104,07 | 106,18 | 108,43 | 103,11 | 96,00 | 99,36 |
| Comunidad Valenciana | 100,00 | 104,50 | 106,59 | 106,70 | 97,01 | 89,33 | 89,86 |
| País Vasco | 100,00 | 103,63 | 106,24 | 107,92 | 104,53 | 97,02 | 88,23 |
| La Rioja | 100,00 | 105,22 | 106,65 | 109,27 | 104,10 | 98,06 | 99,79 |
| Ceuta | 100,00 | 103,77 | 107,21 | 111,17 | 105,47 | 98,85 | 100,59 |
| Melilla | 100,00 | 105,44 | 109,36 | 112,29 | 103,93 | 99,71 | 98,27 |
| NACIONAL | 100,00 | 104,94 | 107,60 | 109,34 | 102,67 | 96,58 | 97,59 |

Tabla 24. Números índice, base=2008, del número de recetas facturadas del Sistema Nacional de Salud

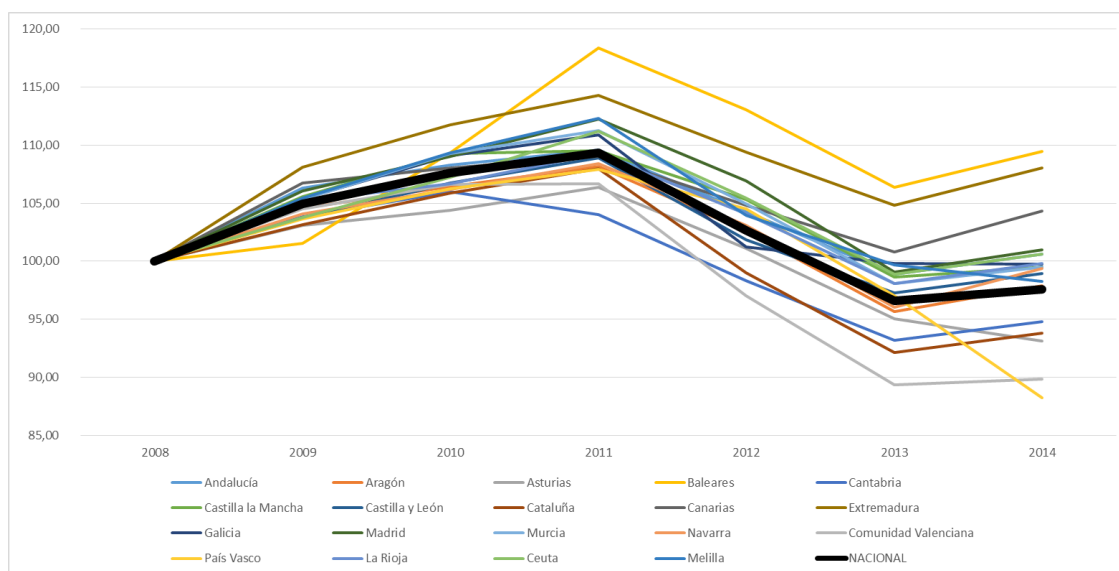


Figura 25. Números índice, base=2008, del número de recetas facturadas del Sistema Nacional de Salud

El comportamiento de la serie de recetas facturadas en lo que se refiere a la cuota que representa cada Comunidad respecto del total nacional, se ofrece en la Figura 26.

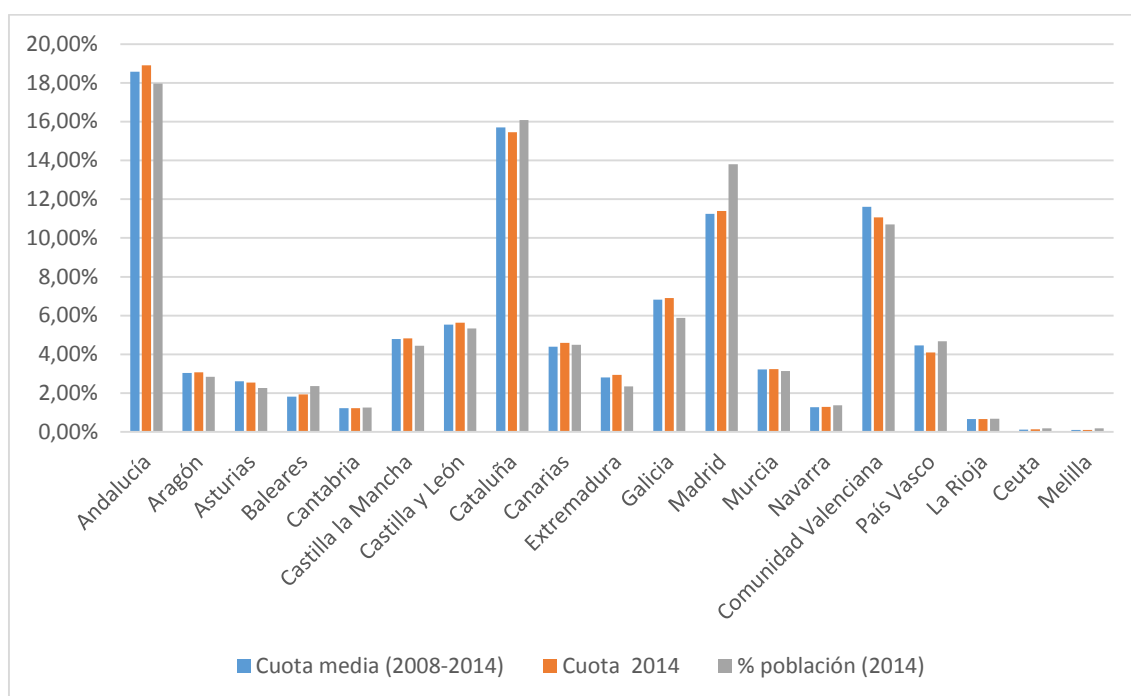


Figura 26. Cuotas de participación sobre el total de recetas facturadas y población (INE 2014)

Respecto a la participación de cada Comunidad sobre el total de recetas facturadas al Sistema Nacional de Salud hemos realizado un análisis de regresión lineal entre cada par de cuotas. Los coeficientes de determinación (R^2) aparecen en la Tabla 25. En ella observamos que entre la cuota media en el periodo 2008-2014 y la cuota en el 2014 se obtiene una relación lineal casi perfecta por lo que podemos considerar que la cuota de cada Comunidad se mantiene constante a lo largo del periodo analizado. Si comparamos el porcentaje poblacional en el 2014 con dichas cuotas, observamos que se obtienen coeficientes de determinación de 0,9816 y 0,9818, respectivamente. En este caso, Madrid con una participación sobre el total de facturas inferior a su porcentaje poblacional, y Galicia y Andalucía superior, hacen que la relación lineal entre las cuotas sea algo menor que en el caso anterior. Aun así, la relación también es muy alta.

| | Cuota media (2008-2014) | Cuota 2014 | % población (2014) |
|------------------------------------|---|---|---|
| Cuota media (2008-2014) | | $R^2 = 0,9986$ La regresión es prácticamente perfecta | $R^2 = 0,9816$ Madrid por debajo de lo esperado Galicia y Andalucía por encima de lo esperado |
| Cuota 2014 | $R^2 = 0,9986$ La regresión es prácticamente perfecta | | $R^2 = 0,9818$ Madrid por debajo de lo esperado Galicia y Andalucía por encima de lo esperado |
| % población (2014) | $R^2 = 0,9816$ Madrid por debajo de lo esperado Galicia y Andalucía por encima de lo esperado | $R^2 = 0,9818$ Madrid por debajo de lo esperado Galicia y Andalucía por encima de lo esperado | |

Tabla 25. Regresión lineal entre las cuotas de participación sobre el total de recetas facturadas y población

| COMUNIDAD | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | TVMA (2003-2014) | Gasto medio (2008-2014) | Cuota media (2008-2014) | Cuota media 2014 |
|----------------------|----------------|----------------|----------------|----------------|---------------|---------------|---------------|---------------------|----------------------------|----------------------------|---------------------|
| Andalucía | 1.955.948.101 | 2.060.988.434 | 2.004.670.059 | 1.823.579.328 | 1.665.242.090 | 1.596.547.350 | 1.636.425.637 | -2,93% | 1.820.485.857 | 16,74% | 17,48% |
| Aragón | 391.319.328 | 409.670.254 | 391.534.064 | 350.535.724 | 309.137.339 | 287.272.191 | 291.978.802 | -4,76% | 347.349.672 | 3,19% | 3,12% |
| Asturias | 338.516.086 | 352.792.046 | 343.741.776 | 317.137.548 | 268.437.256 | 246.386.884 | 252.883.438 | -4,74% | 302.842.148 | 2,78% | 2,70% |
| Baleares | 211.408.067 | 222.875.692 | 219.697.060 | 204.456.285 | 181.461.421 | 172.590.720 | 179.296.813 | -2,71% | 198.826.580 | 1,83% | 1,92% |
| Cantabria | 153.839.717 | 160.330.254 | 152.456.479 | 141.379.553 | 127.561.035 | 119.876.163 | 123.577.212 | -3,58% | 139.860.059 | 1,29% | 1,32% |
| Castilla la Mancha | 583.970.067 | 614.329.118 | 596.484.069 | 539.074.954 | 488.384.411 | 444.425.750 | 451.258.727 | -4,21% | 531.132.442 | 4,88% | 4,82% |
| Castilla y León | 704.476.289 | 741.314.884 | 725.240.336 | 664.273.905 | 561.847.592 | 527.084.123 | 541.280.631 | -4,30% | 637.931.109 | 5,87% | 5,78% |
| Cataluña | 1.841.167.656 | 1.884.515.634 | 1.842.480.622 | 1.682.548.317 | 1.446.250.825 | 1.315.134.931 | 1.343.425.656 | -5,12% | 1.622.217.663 | 14,92% | 14,35% |
| Canarias | 531.575.766 | 569.908.755 | 536.311.937 | 487.444.274 | 421.418.231 | 407.675.137 | 422.903.482 | -3,74% | 482.462.512 | 4,44% | 4,52% |
| Extremadura | 329.427.467 | 358.607.670 | 356.924.545 | 329.405.896 | 288.275.225 | 278.036.095 | 294.596.865 | -1,85% | 319.324.823 | 2,94% | 3,15% |
| Galicia | 869.440.884 | 924.703.358 | 931.283.873 | 829.194.308 | 710.434.191 | 692.758.500 | 681.418.620 | -3,98% | 805.604.819 | 7,41% | 7,28% |
| Madrid | 1.253.153.228 | 1.318.467.244 | 1.288.457.667 | 1.185.733.780 | 1.057.719.802 | 1.003.319.890 | 1.057.406.669 | -2,79% | 1.166.322.611 | 10,72% | 11,30% |
| Murcia | 416.091.706 | 412.286.893 | 420.249.109 | 385.709.194 | 333.921.229 | 313.521.486 | 316.510.619 | -4,46% | 371.184.319 | 3,41% | 3,38% |
| Navarra | 160.142.388 | 166.729.250 | 161.140.260 | 148.267.156 | 128.388.099 | 118.851.809 | 123.723.254 | -4,21% | 143.891.745 | 1,32% | 1,32% |
| Comunidad Valenciana | 1.557.276.381 | 1.604.870.225 | 1.543.952.639 | 1.406.683.620 | 1.199.797.369 | 1.105.948.082 | 1.104.025.534 | -5,57% | 1.360.364.836 | 12,51% | 11,80% |
| País Vasco | 563.037.292 | 586.667.158 | 578.019.562 | 534.209.181 | 490.888.796 | 467.160.037 | 449.427.543 | -3,69% | 524.201.367 | 4,82% | 4,80% |
| La Rioja | 84.228.110 | 89.683.920 | 87.753.275 | 79.272.421 | 67.883.703 | 63.512.237 | 65.134.256 | -4,19% | 76.781.132 | 0,71% | 0,70% |
| Ceuta | 14.347.844 | 14.908.637 | 14.948.650 | 14.399.972 | 13.104.827 | 12.437.978 | 13.031.446 | -1,59% | 13.882.765 | 0,13% | 0,14% |
| Melilla | 11.589.069 | 12.043.485 | 12.337.165 | 12.096.560 | 10.779.928 | 10.710.382 | 11.670.180 | 0,12% | 11.603.824 | 0,11% | 0,12% |
| NACIONAL | 11.970.955.447 | 12.505.692.910 | 12.207.683.146 | 11.135.401.976 | 9.770.933.367 | 9.183.249.747 | 9.359.975.385 | -4,02% | 10.876.270.283 | 100,00% | 100,00% |

Tabla 26. Gasto farmacéutico a través de receta oficial del Sistema Nacional de Salud en el periodo 2008-2014. Tasas de variación y reparto según las CCAA

Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad y elaboración propia

En la Tabla 26 podemos observar el comportamiento global del gasto farmacéutico a través de recetas oficiales. En ella se muestra como la Comunidad Valenciana es la que

menor TVMA presenta con un valor de -5,57%; es decir, que en el periodo 2008-2014 ha experimentado un decrecimiento anual medio acumulado del 5,57% del gasto farmacéutico. Le siguen Cataluña, Aragón y Asturias. En el otro extremo aparece Melilla, que es la única que ha incrementado el gasto farmacéutico a través de recetas (+0,12%). Ceuta (-1,59%) y Extremadura (-1,85%) son las que menos disminuyen su gasto en el periodo. En la Figura 27 se aprecia el comportamiento del gasto farmacéutico a través de recetas oficiales. En dicha figura observamos como nueve Comunidades están por debajo de la media nacional (recordemos que respecto al número de facturas eran solo cinco) y sólo Melilla no bajó su gasto. La disminución a nivel nacional ha sido de un 4,02% de media anual acumulado, en el periodo 2008-2014.

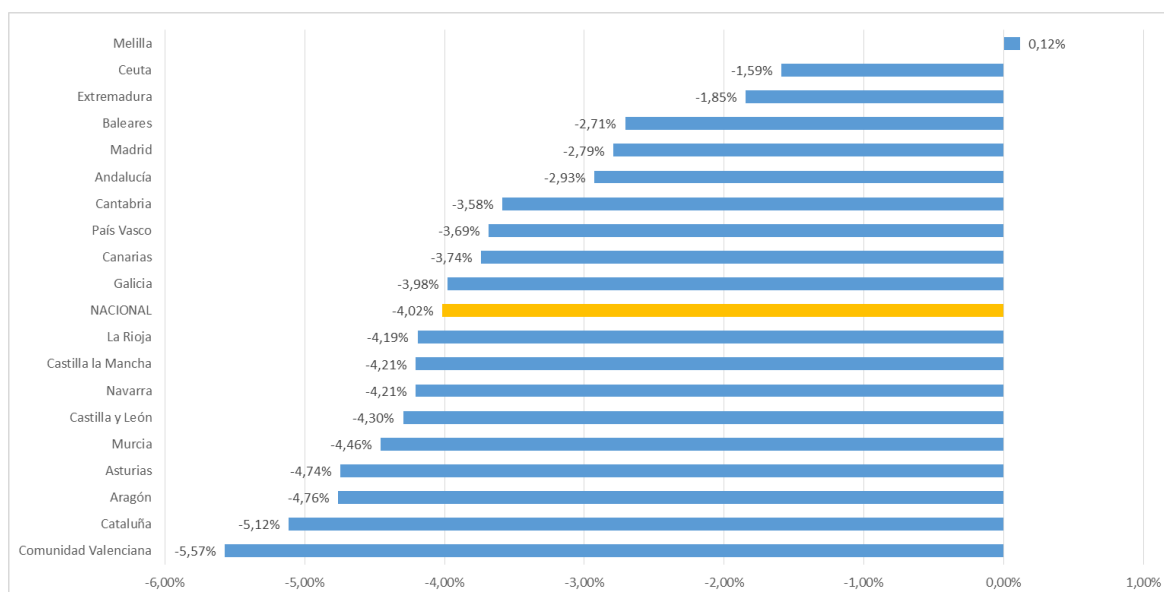


Figura 27. TVMA del gasto farmacéutico a través de receta oficial del Sistema Nacional de Salud en el periodo 2008-2014

De igual forma que en el caso anterior, mediremos la evolución con respecto al año 2008 a través de los números índices simples con base 2008. Dichos índices aparecen en la Tabla 27 y en la Figura 28. En este caso se observa que desde el 2010 cada Comunidad comienza a disminuir su gasto farmacéutico y a partir de entonces cada vez de forma más acusada. Aquí vemos, fielmente, el resultado de la aplicación de las políticas de

contención del gasto farmacéutico público llevadas a cabo por las todas las Comunidades.

| COMUNIDAD | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|----------------------|---------------|---------------|---------------|--------------|--------------|--------------|--------------|
| Andalucía | 100,00 | 105,37 | 102,49 | 93,23 | 85,14 | 81,63 | 83,66 |
| Aragón | 100,00 | 104,69 | 100,05 | 89,58 | 79,00 | 73,41 | 74,61 |
| Asturias | 100,00 | 104,22 | 101,54 | 93,68 | 79,30 | 72,78 | 74,70 |
| Baleares | 100,00 | 105,42 | 103,92 | 96,71 | 85,83 | 81,64 | 84,81 |
| Cantabria | 100,00 | 104,22 | 99,10 | 91,90 | 82,92 | 77,92 | 80,33 |
| Castilla la Mancha | 100,00 | 105,20 | 102,14 | 92,31 | 83,63 | 76,10 | 77,27 |
| Castilla y León | 100,00 | 105,23 | 102,95 | 94,29 | 79,75 | 74,82 | 76,83 |
| Cataluña | 100,00 | 102,35 | 100,07 | 91,38 | 78,55 | 71,43 | 72,97 |
| Canarias | 100,00 | 107,21 | 100,89 | 91,70 | 79,28 | 76,69 | 79,56 |
| Extremadura | 100,00 | 108,86 | 108,35 | 99,99 | 87,51 | 84,40 | 89,43 |
| Galicia | 100,00 | 106,36 | 107,11 | 95,37 | 81,71 | 79,68 | 78,37 |
| Madrid | 100,00 | 105,21 | 102,82 | 94,62 | 84,40 | 80,06 | 84,38 |
| Murcia | 100,00 | 99,09 | 101,00 | 92,70 | 80,25 | 75,35 | 76,07 |
| Navarra | 100,00 | 104,11 | 100,62 | 92,58 | 80,17 | 74,22 | 77,26 |
| Comunidad Valenciana | 100,00 | 103,06 | 99,14 | 90,33 | 77,04 | 71,02 | 70,89 |
| País Vasco | 100,00 | 104,20 | 102,66 | 94,88 | 87,19 | 82,97 | 79,82 |
| La Rioja | 100,00 | 106,48 | 104,19 | 94,12 | 80,60 | 75,41 | 77,33 |
| Ceuta | 100,00 | 103,91 | 104,19 | 100,36 | 91,34 | 86,69 | 90,83 |
| Melilla | 100,00 | 103,92 | 106,46 | 104,38 | 93,02 | 92,42 | 100,70 |
| NACIONAL | 100,00 | 104,47 | 101,98 | 93,02 | 81,62 | 76,71 | 78,19 |

Tabla 27. Números índice, base=2008, del gasto farmacéutico a través de receta oficial del Sistema Nacional de Salud

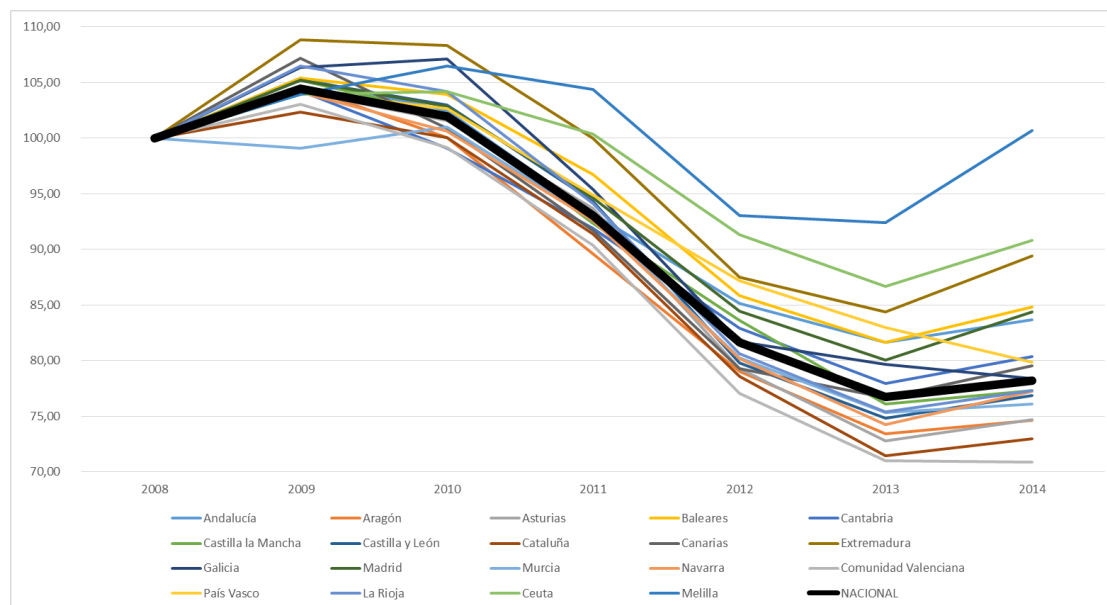


Figura 28. Números índice, base=2008, del gasto farmacéutico a través de receta oficial del Sistema Nacional de Salud

El comportamiento del gasto farmacéutico a través de receta oficial respecto de la cuota que representa cada Comunidad sobre el total nacional, se observa en la Figura 29.

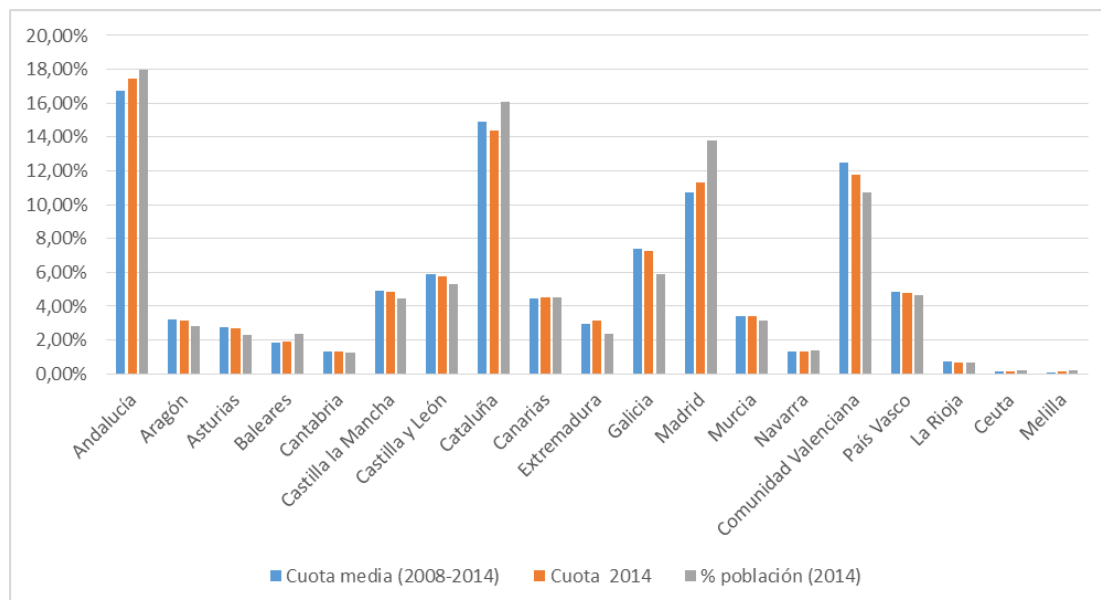


Figura 29. Cuotas de participación sobre el total del gasto farmacéutico en recetas y población (INE 2014)

Respecto a la participación de cada Comunidad sobre el total del gasto farmacéutico a través de receta oficial hemos repetido el análisis de regresión lineal que ya hicimos con el indicador del número de recetas. En la Tabla 28 observamos que entre la cuota media y la cuota en el 2014 se obtiene una relación lineal casi sin error, así que podemos considerar que la cuota de cada Comunidad se mantiene constante a lo largo del periodo analizado. Si comparamos el porcentaje poblacional en el 2014 con dichas cuotas, observamos que se obtienen coeficientes de determinación de 0,9658 y 0,9764, respectivamente. En este caso Madrid y Cataluña con una participación sobre el total inferior a su porcentaje poblacional, y Galicia y la Comunidad Valenciana superior, hacen que la relación lineal entre las cuotas sea algo menor que en el caso anterior. Aun así, la relación también es muy alta.

| | Cuota media (2008-2014) | Cuota 2014 | % población (2014) |
|----------------------------|---|---|---|
| Cuota media (2008-2014) | R ² = 0,9960 La regresión es prácticamente perfecta. Andalucía y Madrid han disminuido el gasto en el periodo | R ² = 0,9960 | R ² = 0,9658 |
| | | Madrid por debajo de lo esperado | Galicia y C. Valenciana por encima de lo esperado |
| | | | |
| Cuota 2014 | R ² = 0,9960 La regresión es prácticamente perfecta. Andalucía y Madrid han disminuido el gasto en el periodo | | R ² = 0,9764 |
| | | | Madrid y Cataluña por debajo de lo esperado |
| | | | Galicia y C. Valenciana por encima de lo esperado |
| % población (2014) | R ² = 0,9658 | R ² = 0,9764 | |
| | Madrid por debajo de lo esperado | Madrid y Cataluña por debajo de lo esperado | |
| | Galicia y C. Valenciana por encima de lo esperado | Galicia y C. Valenciana por encima de lo esperado | |

Tabla 28. Regresión lineal entre las cuotas de participación sobre el total del gasto en recetas y población

Por último, vamos a analizar el indicador combinado “gasto medio por receta” cuyos datos aparecen en la Tabla 29. En dicha tabla observamos que el gasto medio por receta ha disminuido en España en el periodo 2008-2014 un 3,63 % por receta anualmente, de forma acumulada.

| COMUNIDAD | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | TVMA (2008-2014) | Gasto medio por receta (2008-2014) |
|----------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------------|--|
| Andalucía | 11,98 | 11,87 | 11,34 | 10,19 | 9,79 | 9,89 | 9,96 | -3,03% | 10,72 |
| Aragón | 14,39 | 14,40 | 13,53 | 11,92 | 11,08 | 11,04 | 10,96 | -4,43% | 12,47 |
| Asturias | 14,23 | 14,39 | 13,84 | 12,53 | 11,16 | 10,90 | 11,42 | -3,61% | 12,64 |
| Baleares | 13,82 | 14,34 | 13,13 | 11,29 | 10,49 | 10,61 | 10,71 | -4,16% | 12,06 |
| Cantabria | 13,74 | 13,79 | 12,84 | 12,14 | 11,59 | 11,49 | 11,64 | -2,72% | 12,46 |
| Castilla la Mancha | 13,89 | 13,84 | 12,98 | 11,71 | 11,03 | 10,72 | 10,79 | -4,12% | 12,14 |
| Castilla y León | 14,26 | 14,47 | 13,75 | 12,34 | 11,16 | 10,97 | 11,07 | -4,12% | 12,58 |
| Cataluña | 12,87 | 12,77 | 12,16 | 10,89 | 10,21 | 9,98 | 10,01 | -4,10% | 11,27 |
| Canarias | 13,90 | 13,96 | 12,98 | 11,69 | 10,51 | 10,58 | 10,60 | -4,42% | 12,03 |
| Extremadura | 13,90 | 14,00 | 13,48 | 12,17 | 11,12 | 11,19 | 11,51 | -3,10% | 12,48 |
| Galicia | 14,46 | 14,59 | 14,20 | 12,44 | 11,67 | 11,55 | 11,36 | -3,94% | 12,90 |
| Madrid | 12,79 | 12,69 | 12,06 | 10,78 | 10,10 | 10,34 | 10,69 | -2,95% | 11,35 |
| Murcia | 14,73 | 13,86 | 13,62 | 12,27 | 11,27 | 11,31 | 11,26 | -4,38% | 12,62 |
| Navarra | 14,12 | 14,13 | 13,38 | 12,06 | 10,98 | 10,92 | 10,98 | -4,11% | 12,36 |
| Comunidad Valenciana | 14,56 | 14,36 | 13,54 | 12,32 | 11,56 | 11,57 | 11,49 | -3,87% | 12,77 |
| País Vasco | 13,95 | 14,03 | 13,48 | 12,27 | 11,64 | 11,93 | 12,62 | -1,65% | 12,85 |
| La Rioja | 14,46 | 14,63 | 14,12 | 12,45 | 11,19 | 11,12 | 11,20 | -4,16% | 12,74 |
| Ceuta | 13,00 | 13,02 | 12,63 | 11,73 | 11,26 | 11,40 | 11,74 | -1,69% | 12,11 |
| Melilla | 12,40 | 12,22 | 12,07 | 11,52 | 11,10 | 11,49 | 12,70 | 0,41% | 11,93 |
| NACIONAL | 13,45 | 13,39 | 12,75 | 11,44 | 10,69 | 10,68 | 10,78 | -3,63% | 11,88 |

Tabla 29. Gasto medio por receta facturada del Sistema Nacional de Salud en el periodo 2008-2014. Tasas de variación y reparto según las CCAA

Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad y elaboración propia

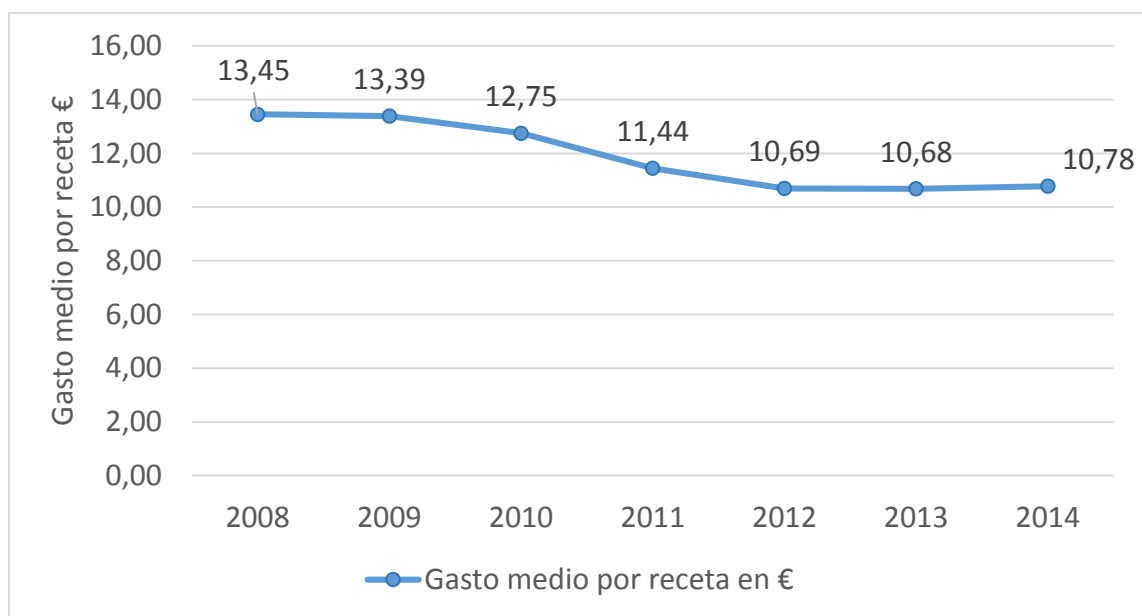


Figura 30. Evolución del gasto medio por receta facturada del Sistema Nacional de Salud en el periodo 2008-2014

En la Figura 30 vemos que el gasto medio por receta tiene un comportamiento decreciente de forma suave, pasando de 13,45 € en el 2008 a 10,78 € en el 2014.

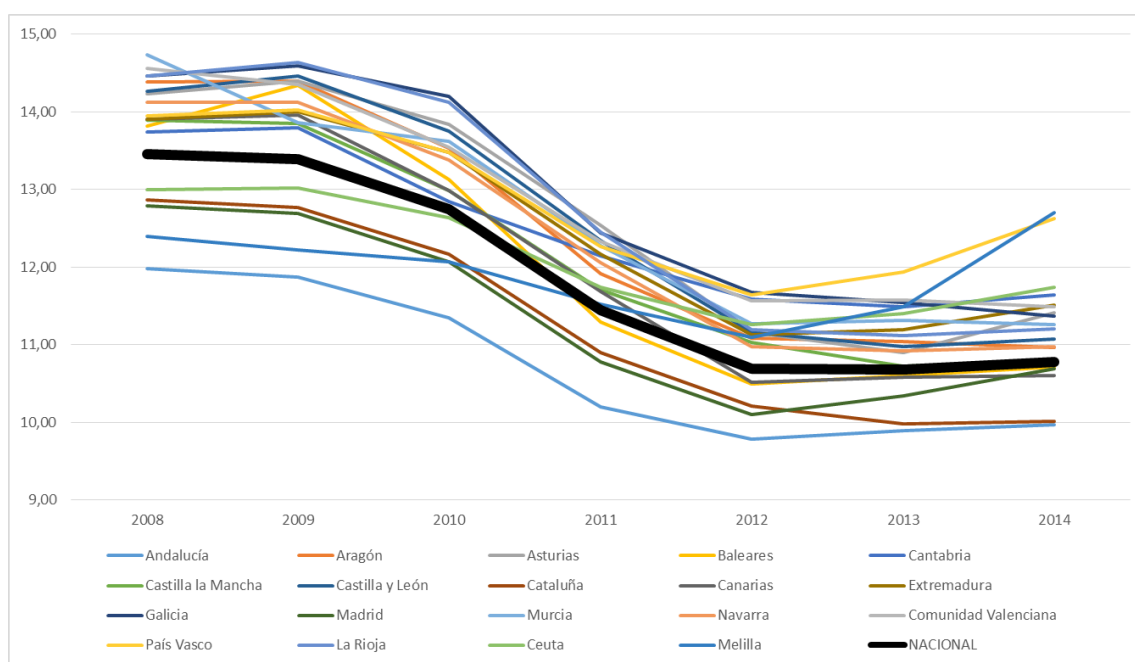


Figura 31. Comparación del gasto medio por receta por Comunidad Autónoma

En la Figura 31 observamos como a nivel de Comunidad el comportamiento es homogéneo y ha evolucionado de forma similar en toda España.

Un estudio realizado (60) para Farmaindustria señala que el crecimiento del gasto farmacéutico está relacionado positivamente con el gasto sanitario, el gasto público y la proporción de personas mayores de 65 años. El estudio también concluye que no se distingue en España un patrón de evolución del consumo farmacéutico diferente al de otros países europeos. Sin embargo, este modelo no incluye variables que son relevantes en la explicación del gasto, como por ejemplo el nivel de precios, el volumen de recetas prescritas, la proporción que los genéricos suponen sobre el mercado o el nivel de copago establecido, entre otras. A priori, parece lógico pensar que el gasto farmacéutico estaría relacionado con el nivel del PIB o el gasto sanitario, pero esas variables no explicarían la magnitud de su crecimiento.

El gasto de los países de la OCDE en medicamentos, ya sean de prescripción o sin receta, representa en promedio un 1,5% del PIB. El 85% corresponde a medicamentos de prescripción, y un 15% a la venta de medicamentos sin receta (61).

1.2.5.2 Mercado del autocuidado de la salud

Según la Asociación para el Autocuidado de la Salud (anefp) (64) el mercado del autocuidado presenta la siguiente estructura en cuanto a productos:

| MERCADO DEL AUTOCUIDADO | |
|---|---|
| Medicamentos de autocuidado | Productos de autocuidado |
| Medicamentos no sujetos a prescripción y no financiados | Productos sanitarios no financiados |
| Medicamentos a base de plantas | Cosmética y dermofarmacia |
| Medicamentos homeopáticos | Productos alimenticios (complementos alimenticios, dietéticos y nutrición infantil) |

Tabla 30. Estructura del mercado del autocuidado

A su vez, el mercado está dividido en cuatro segmentos representados en la Figura 32.



Figura 32. Segmentos en el mercado del autocuidado

En el segmento de OTC (*Over The Counter*) se engloban los medicamentos sin receta no financiados y productos de consumo de similar posicionamiento. En este segmento existen tres categorías dependiendo del tipo de registros y de su situación de reembolso:

- Semi-éticos: son los medicamentos que pueden dispensarse tanto con receta como sin receta.
- Consumo: productos sin registro de medicamento como suplementos alimenticios o plantas medicinales.
- Medicamentos publicitarios: medicamentos que se dispensan sin prescripción y no están reembolsados.

En el segmento de PEC se agrupan los productos para el cuidado personal. En este segmento existen dos categorías:

- Higiene oral.
- Dermoestética.

En el segmento de PAC quedan incluidos los productos para el cuidado del paciente existiendo en él dos categorías:

- Incontinencia.
- Accesorios sanitarios y otros.

Finalmente, en el segmento de NUT se incluyen los productos de nutrición existiendo tres categorías:

- Dieta y otros.
- Enteral.
- Infantil.

Las cuotas de mercado (65) según PVP en el año 2014 de los segmentos pueden verse en la Figura 33. El desglose de la cuota según sus respectivas categorías se aprecia en la Figura 34.

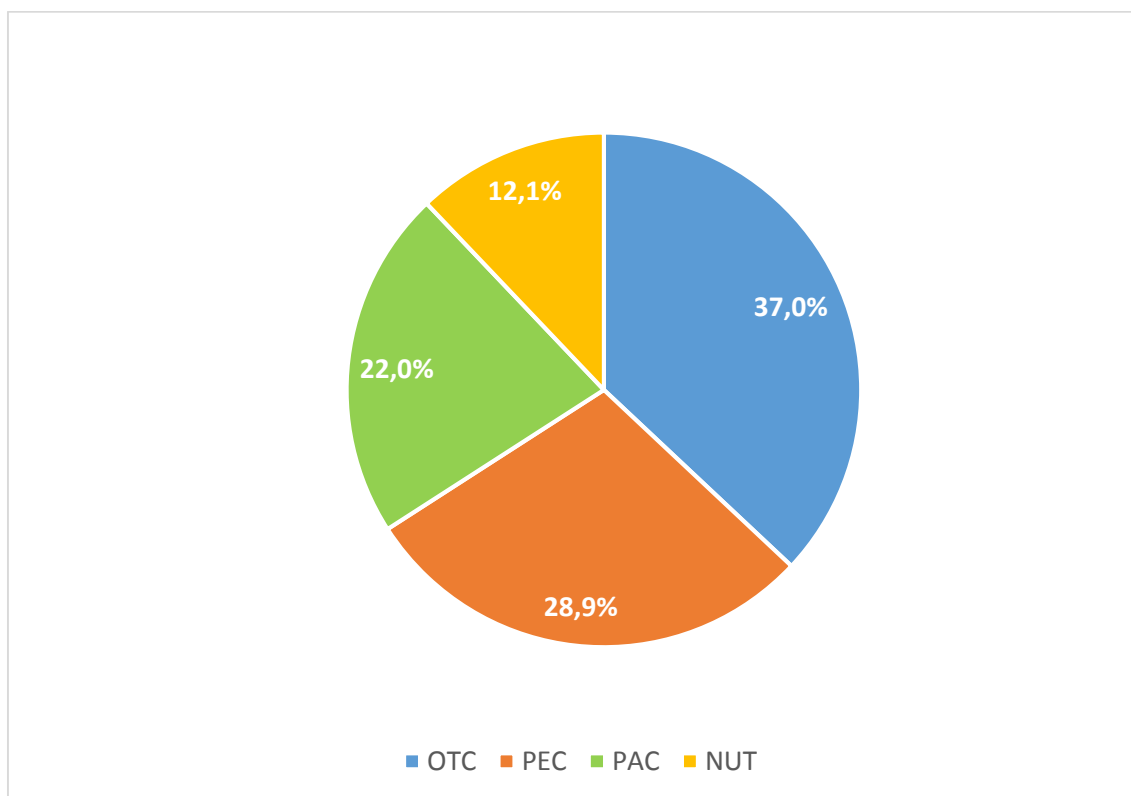


Figura 33. Distribución de los segmentos del mercado de Consumer Health (%€ PVP; 2014)

Fuente: Análisis IMS Health, Sell-out (Combi TAM 12/2014) Muestra 3.449 farmacias

Podemos observar en la Figura 33 que el segmento de los medicamentos sin receta no financiados tiene la mayor cuota de mercado pero sin embargo si observamos la Figura 34 podemos advertir que la mayor cuota de mercado a nivel de categoría la presenta la dermoestética que representa el 80,62% de su segmento.

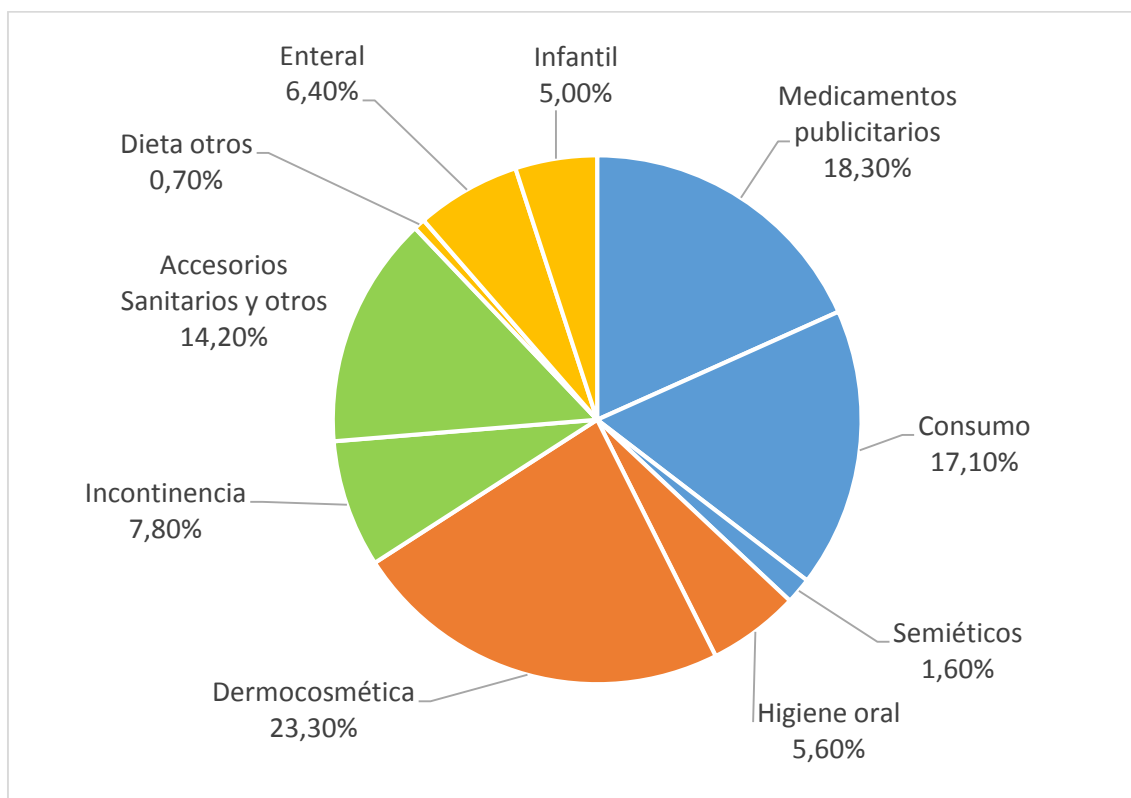


Figura 34. Distribución de las segmentos y categorías del mercado de Consumer Health (%€ PVP; 2014)

Fuente: Análisis IMS Health, Sell-out (Combi TAM 12/2014) Muestra 3.449 farmacias

Hay que destacar que la mayoría de los productos de autocuidado de la salud (complementos alimenticios, higiene oral, accesorios sanitarios, nutrición infantil, productos adelgazantes) se pueden vender en otros canales como son las parafarmacias y los supermercados. Además, tienen el precio libre y se pueden anunciar en cualquier tipo de soporte. En este mercado la oficina de farmacia se enfrenta a una competencia diferente a la que tiene en el mercado de prescripción. En este caso, la competencia está habituada a luchar más en un mercado no regulado y utiliza técnicas de marketing que no tienen cabida en el mercado regulado.

Con respecto al comportamiento de este mercado podemos ver su evolución, en valores y unidades, en la Tabla 31.

| | Valores | Variación % | Unidades | Variación % |
|-------------|---------|-------------|----------|-------------|
| 2010 | 5.120 | | 601 | |
| 2011 | 4.967 | -2,99% | 566 | -5,82% |
| 2012 | 4.776 | -3,85% | 526 | -7,07% |
| 2013 | 4.830 | 1,13% | 503 | -4,37% |
| 2014 | 5.050 | 4,55% | 509 | 1,19% |

Tabla 31. Evolución del mercado de autocuidado en valores (PVP €) y unidades. 2010-2014

Fuente: COMBISP (Sell out); IMS Health Análisis

El comportamiento de la serie, Figura 35, es a la baja desde el 2010 hasta el 2013, año en el cual se produce el primer aumento de los últimos años. Como veremos más adelante, no todos los segmentos del mercado ayudan a este repunte su actividad.

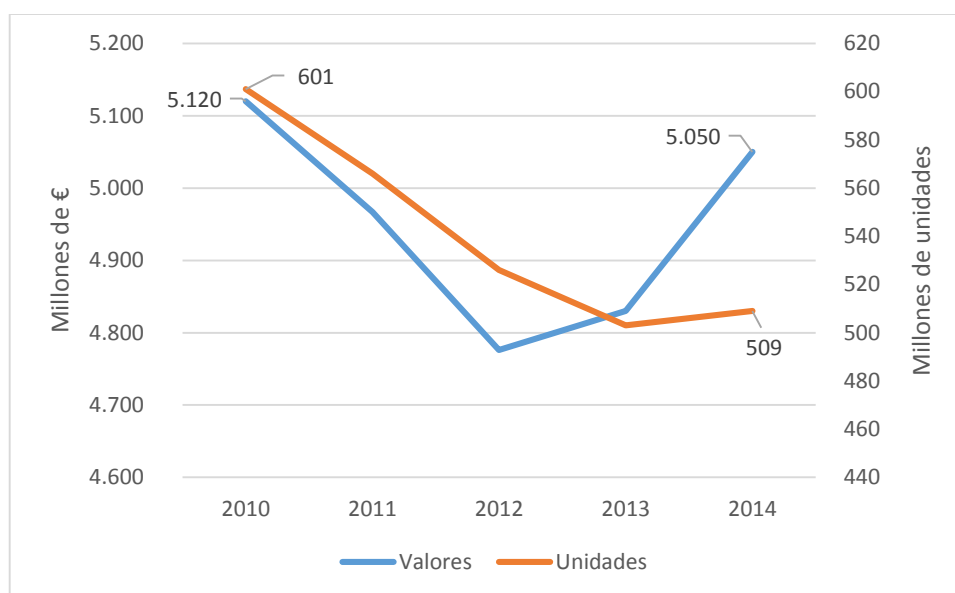


Figura 35. Evolución del mercado de autocuidado en valores (PVP €) y unidades. 2010-2014

Fuente: COMBISP (Sell out); IMS Health Análisis

En la Tabla 32 se desglosa el comportamiento del mercado según los cuatros segmentos que lo conforman. En dicha tabla se observa que en el caso de los segmentos OTC y PEC se han producido tasas de variación interanual positivas respecto al volumen en € en los dos últimos años. En los segmentos PAC y NUT el crecimiento se produce solo en el 2014.

| Segmento | | 2010 | 2011 | 2012 | 2013 | 2014 |
|----------|-------------|----------|----------|----------|----------|----------|
| OTC | Valores | 1.770,00 | 1.732,00 | 1.641,00 | 1.738,00 | 1.867,00 |
| | Variación % | | -2,15% | -5,25% | 5,91% | 7,42% |
| | Unidades | 293,00 | 276,00 | 251,00 | 241,00 | 245,00 |
| | Variación % | | -5,80% | -9,06% | -3,98% | 1,66% |
| PEC | Valores | 1.476,00 | 1.426,00 | 1.394,00 | 1.409,00 | 1.461,00 |
| | Variación % | | -3,39% | -2,24% | 1,08% | 3,69% |
| | Unidades | 159,00 | 148,00 | 142,00 | 138,00 | 139,00 |
| | Variación % | | -6,92% | -4,05% | -2,82% | 0,72% |
| PAC | Valores | 1.190,00 | 1.130,00 | 1.095,00 | 1.081,00 | 1.114,00 |
| | Variación % | | -5,04% | -3,10% | -1,28% | 3,05% |
| | Unidades | 93,00 | 91,00 | 86,00 | 83,00 | 86,00 |
| | Variación % | | -2,15% | -5,49% | -3,49% | 3,61% |
| NUT | Valores | 684,00 | 679,00 | 646,00 | 602,00 | 608,00 |
| | Variación % | | -0,73% | -4,86% | -6,81% | 1,00% |
| | Unidades | 56,00 | 51,00 | 47,00 | 41,00 | 39,00 |
| | Variación % | | -8,93% | -7,84% | -12,77% | -4,88% |

Tabla 32. Evolución del mercado de autocuidados, según segmentos, en valores (PVP €) y unidades. 2010-2014

Fuente: COMBISP (Sell out); IMS Health Análisis

Si analizamos el comportamiento del mercado con respecto al número de unidades vendidas, observamos que la supuesta recuperación no se ve refrendada. El segmento que mejor recuperación ha experimentado en el 2014 ha sido el de PAC. En el segmento NUT el comportamiento es claramente a la baja, si bien, en el 2014 se da la menor caída de la serie en el periodo 2010-2014.

En la Figura 36 se aprecia el comportamiento del volumen del mercado en millones de € a precios PVP por segmentos y se aprecia la tasa de variación interanual del total del mercado. Podemos advertir como la cuota de cada segmento se mantiene a lo largo de los años y que el crecimiento final es debido fundamentalmente al segmento OTC.

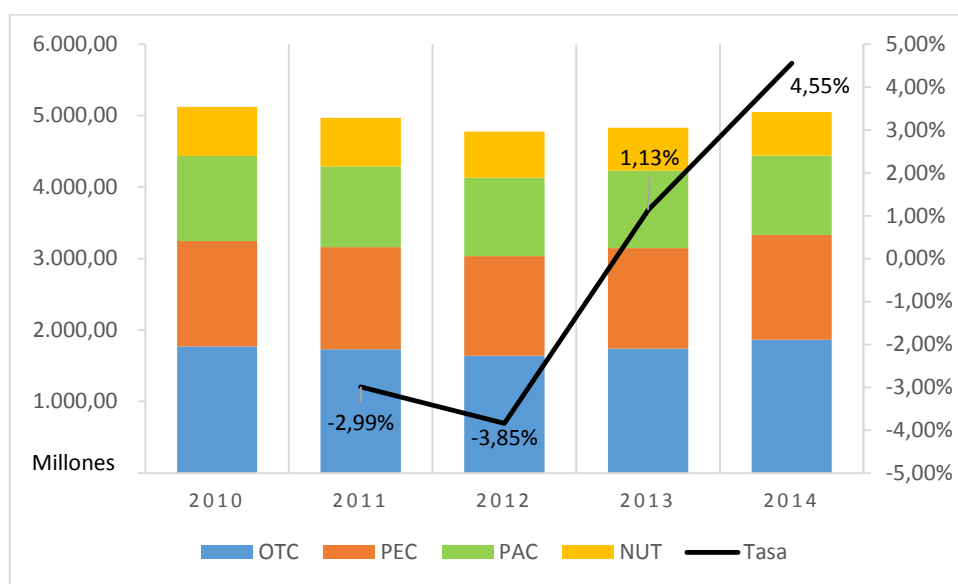


Figura 36. Evolución del mercado de autocuidados, según segmentos, en valores (PVP €). 2010-2014

Fuente: COMBISP (Sell out); IMS Health Análisis

En la Figura 37 se muestra el comportamiento del mercado en millones de unidades vendidas. Podemos observar que la evolución del mercado es negativa en cuanto a unidades vendidas. Por lo tanto, si resulta que el mercado medido en valores crece y medido en unidades decrece, los precios de los productos de este mercado han experimentado una subida lo suficientemente importante para que el descenso de unidades no haya podido con el aumento del volumen en € del mercado.

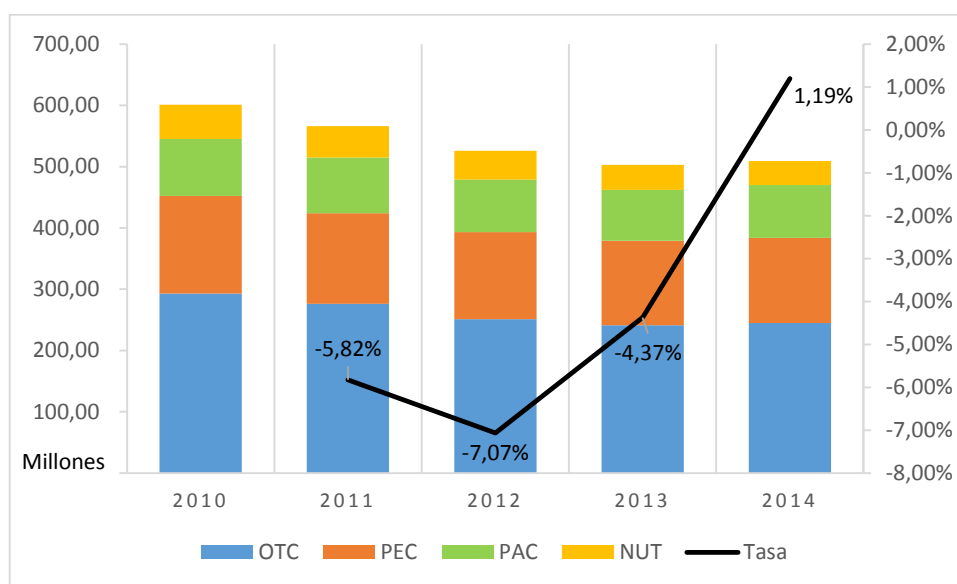


Figura 37. Evolución del mercado de autocuidados, según segmentos, en millones de unidades. 2010-2014

Fuente: COMBISP (Sell out); IMS Health Análisis

Por último, estudiaremos las características del mercado de los medicamentos publicitarios por el peso que tienen en las oficinas de farmacia y porque es en ella el único establecimiento en el que se pueden comprar. En el sector, estos medicamentos son conocidos como Especialidades Farmacéuticas Publicitarias (EFP), no son financiados por el Sistema Nacional de Salud y su precio es libre.

Como pudimos ver antes, aunque en lo referente al volumen ha tenido un comportamiento al alza en los dos últimos años, respecto al número de unidades ocurre lo contrario. El aumento del volumen ha sido provocado por el incremento del precio medio por parte de los laboratorios (64).

Según Javier Navarro (64), el Ministerio de Sanidad debería potenciar este mercado ya que el tratamiento de síntomas menores con medicamentos publicitarios permite:

- Reducir el coste de los medicamentos financiados a su cargo.
- Disminuir la presión sobre las sobresaturadas consultas de atención primaria, con el consiguiente ahorro indirecto.
- Liberar recursos para otras necesidades sanitarias.

En el mercado del autocuidado de la salud y en concreto en la automedicación, el farmacéutico adquiere un papel fundamental. De hecho, el farmacéutico es en la mayoría de las ocasiones el único profesional sanitario testigo y observador de la automedicación y esto hace que su rol sea fundamental para (65):

- Prevenir el mal uso de medicamentos.
- Dar consejos correctos y adecuados.
- Seleccionar, supervisar y/o indicar un medicamento OTC.
- Detectar y notificar reacciones adversas a medicamentos, que están fuera del control del médico.

El cambio de foco del médico hacia el farmacéutico en este mercado hace que los laboratorios actúen sobre la capacidad de indicación del farmacéutico, además de sobre el paciente con campañas publicitarias, y esto provoca algún desajuste en la relación que los agentes guardan el mercado de la prescripción. De hecho, la distribución de los medicamentos publicitarios puede ser de dos formas (64):

- Venta al mayorista: que puede ser a demanda o forzada a través de los denominados “transfer”, esto es, pedidos que obtienen los vendedores de los laboratorios farmacéuticos pero que son servidos por los mayoristas.
- Venta directa a la oficina de farmacia: los laboratorios sirven directamente pedidos realizados por sus vendedores.

Los medicamentos publicitarios, como su propio nombre indica, son objeto de publicidad destinada al público en general sujeta a la legalidad vigente. Por ley, cuando esta publicidad es emitida en medios audiovisuales, el mensaje debe acabar con la pantalla azul reglamentaria de la Figura 38.

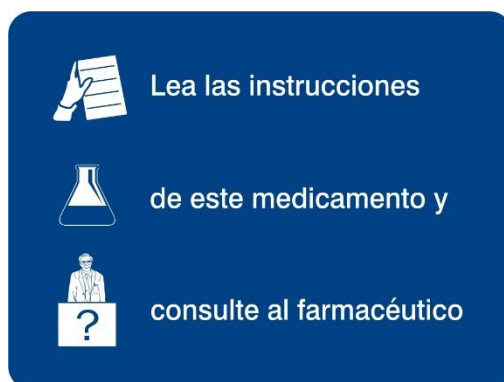


Figura 38. Pantalla azul con la que debe finalizar cualquier publicidad de medicamentos publicitarios en medios audiovisuales

Más allá del reconocimiento por parte de la mayoría de la población de esta pantalla, su imperativo legal refuerza al farmacéutico como centro de atención por parte de los ciudadanos y los laboratorios farmacéuticos para la dispensación de medicamentos publicitarios. Evidentemente, cada uno con un motivo distinto.

Entre los años 2002 y 2007 se observó que un 60% de la población española consumió medicamentos publicitarios, la mayoría aconsejada por el farmacéutico (39%) o basándose en su experiencia personal previa (aproximadamente un 69%) (68).

1.2.6 Situación de las oficinas de farmacia

Desde hace tiempo, especialmente durante las últimas dos décadas, la farmacia española ha venido trabajando para conseguir un modelo que permita la prestación de un servicio acorde con el resto de prestaciones llevadas a cabo por el Sistema Nacional de Salud.

En el mercado regulado de prescripción, a día de hoy el de mayor cuota de facturación en la farmacia española (67), se han desplegado 29 leyes desde 1996 que han moderado, en menor o mayor medida, la tasa de crecimiento de la facturación de las oficinas de farmacia provocando un descenso del margen neto de las farmacias hasta llegar a un 6,9 %, como podemos ver en la Figura 39. Todas estas medidas han tenido y tienen como objetivo la contención del gasto farmacéutico y resolver, en parte, los problemas de la financiación sanitaria (68).

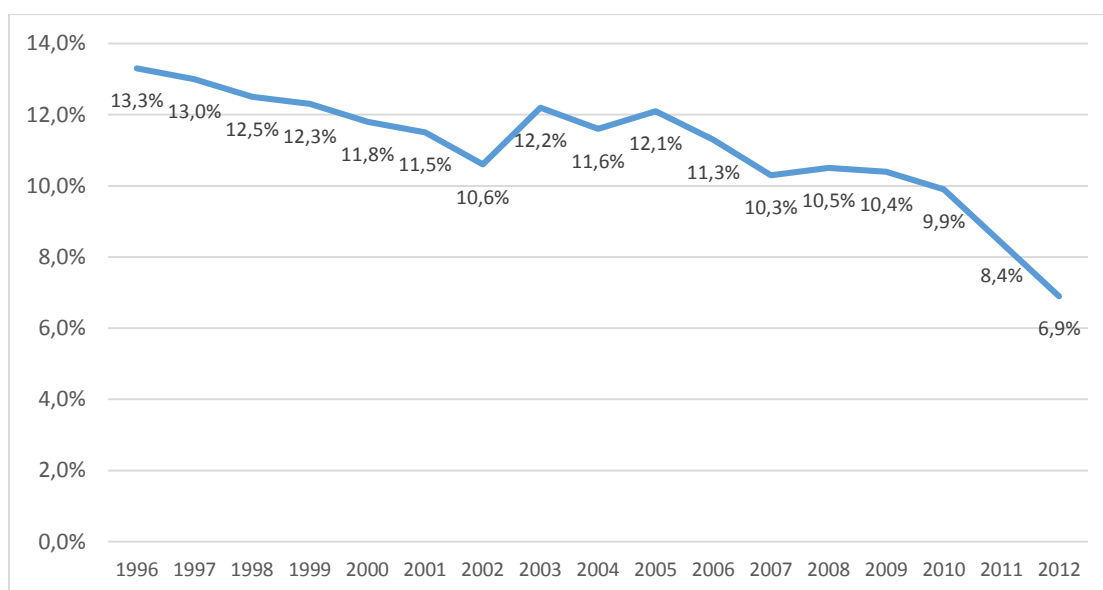


Figura 39. Evolución del margen neto de una Oficina de Farmacia tipo (*) (1996-2012)

(*)Facturación anual de 600.000€ y venta libre del 35%

En lo que respecta al resultado económico y fiscal de las oficinas de farmacia (67), podemos ver en la Tabla 33 un resumen del volumen de facturación anual. En ella, se observan disminuciones importantes de facturación, incrementos del porcentaje de

gastos de personal y una tendencia a la baja de los márgenes de todos los tipos de farmacias.

| Facturación anual (€) | Mix ventas | | Costes laborales | Efecto RDLs | Margen neto antes IRPF |
|------------------------------|------------|--------|--|---|---|
| | SOE | Libre | | | |
| Hasta 300.000 | 71,09% | 28,91% | Incremento al 7,41% sobre Ventas | Nuevo incremento del 11,39% s/2011 | Disminución hasta el 7,21% s/ventas (7,71% en 2.011) |
| 300.000 - 600.000 | 59,36% | 40,64% | Sigue sin bajar del 10% : 10,26 % s/ventas | Se sitúa en por encima de los 7.000 e/año | Nuevo descenso 6,5% s/ventas - 6,87% 2011 |
| 600.000 - 900.000 | 71,29% | 28,71% | Cota máxima serie 2002-2012 : 12,66 % | Cifra promedio > 14.000 €/año | Se sitúa por 2º año consecutivo por debajo del 7% : 6,95% |
| 900.000 - 1.200.000 | 60,20% | 39,80% | Sube al 11,88 % (máximo desde 2003) | Cifra promedio > 33.000 €/año | Nueva bajada al 6,42% s/ventas *6,68% en 2011 |
| 1.200.000 - 2.000.000 | 64,90% | 35,10% | Nuevo incremento al 12,54% s/ventas | Cifra promedio > 70.400 €/año | Descenso al 6,09% s/ventas *6,23% en 2011 |
| Más de 2.000.000 | 47,44% | 52,56% | Se eleva al 15,12 % (Valor máximo del Informe) | Sube hasta más de 195.000 €/año | Bajada al 5,39% s/ventas. |

Tabla 33. Resultado económico-fiscal de las oficinas de farmacia 2013 (BASE: Declaraciones IRPF / Patrimonio 2.012)
Fuente: Informe anual de la Oficinas de Farmacia 2014.ASPIME.

Además, en las cuentas de resultados irrumpen de lleno los costes financieros debido al impago de las administraciones cuya situación en el año 2014 es la que se presenta en la Figura 40.



Figura 40. La situación de los impagos en las Comunidades Autónomas.2014

El Gobierno Autonómico salda la deuda que arrastraba con las farmacias.

- Aragón.
- Baleares.
- Canarias.

CCAA en riesgo y tensiones en el pago de la factura farmacéutica.

- Castilla y León.
- Castilla la Mancha.
- Extremadura.
- Murcia.
- La Rioja.

Cronificación de la deuda.

- Cataluña.
- Comunidad Valenciana.

Tomando como referencia temporal el año 2014 y el informe de Antares Consulting (69), se observa que la farmacia española presenta una situación financiera delicada que no está causada por una única medida, coyuntura o suceso, sino por una acumulación de

todas estas realidades. En el informe se afirma que un 20% de las farmacias españolas se encuentran en una situación financiera desfavorable:

- Un 5% presenta alto riesgo de liquidación o de venderse de forma inadecuada.
- Un 15% tiene dificultades financieras serias:
 - No encuentran fácil acceso a la financiación standard.
 - Están “apalancadas” con los proveedores (distribución).

Se da la circunstancia que las medidas tomadas por la Administración en cuanto a la reducción del gasto farmacéutico como consecuencia de la crisis económica se han centrado fundamentalmente en los medicamentos que se dispensan a través de oficina de farmacia, y muy poco en aquellos que se dispensan o administran en hospitales, teniendo estos un mayor coste (70); por ello, una de las estrategias de los laboratorios farmacéuticos ha sido desplazar la dispensación de sus medicamentos más estratégicos hacia el ámbito hospitalario. De hecho, el incremento del gasto en farmacia hospitalaria que repercute en la cifra global del gasto en medicamentos ha sido muchas veces erróneamente atribuida en exclusiva a la oficina de farmacia (71).

Los cambios del entorno regulatorio español han condicionado la evolución del mercado de prescripción (65) como podemos ver en la Figura 41.

Los Reales Decretos RDL 09/2011 (34) y RDL 16/2012 (8) representan los cambios más radicales en cuanto a las reglas del mercado.

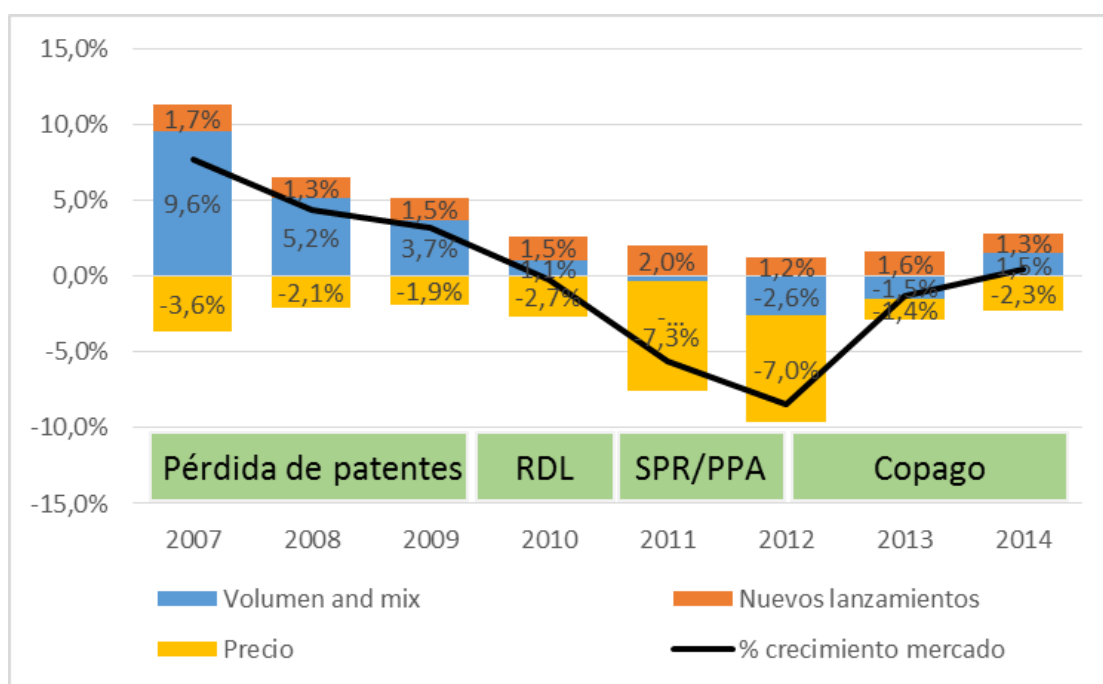


Figura 41. Evolución del mercado de prescripción y cambios regulatorios 2007-2014

Según la consultora *IMS Health* (65) los aspectos más relevantes del mercado de prescripción desde el año 2009 hasta la fecha son los siguientes:

- 2009: reducción de precios por pérdida de patentes.
- 2010-2011: los RDL 04/2010, RDL 08/2010 y RDL 09/2011 convierten la desaceleración en crecimiento negativo dirigido por la reducción precios.
- 2012: RDL 16/2012, efecto negativo del factor volumen.
- 2013: en julio desaparece el efecto negativo en el volumen.
- 2014: se recupera el efecto de contención debido al volumen. El efecto precio queda a la espera de la aplicación de la Orden de Precios de Referencia en septiembre de 2015.

A modo de resumen, podemos destacar que en el mercado de prescripción de las oficinas de farmacia se han producido los siguientes hechos relevantes y sus respectivos efectos negativos para su negocio:

- Pérdidas de patentes.
- Aparición de nuevos RDLs.
- Nuevas órdenes de precios de referencia.

- Aumento del copago de los pacientes.
- Impago por parte de los Servicios Regionales de Salud.
- Aumento de los costes de financiación.
- Incremento de los costes laborales.

Todos estos hechos han provocado los siguientes efectos negativos en las farmacias:

- Empeoramiento de la situación financiera de las farmacias.
- Disminución de los márgenes comerciales.
- Disminución del importe del ticket medio.

En lo que respecta al mercado del autocuidado de la salud, en el 2014 se frena el ritmo de decrecimiento de los últimos 4 años. En 2015 todos los segmentos crecen por encima del 2014 (65):

- Después de una severa desaceleración en 2008 a causa de la crisis financiera, en 2009, el éxito de algunos nuevos lanzamientos y una estacionalidad positiva fueron las razones del impulso del mercado.
- La falta de nuevos lanzamientos relevantes provocó de nuevo el estancamiento del mercado en 2010 y 2011.
- Posteriormente, a mediados de 2011, el impacto de las medidas regulatorias (RDLs) afectó negativamente a todas las categorías con productos reembolsados o que tenían un sustituto reembolsado.
- En los últimos meses de 2013 y los dos primeros trimestres de 2014 el mercado mostró de nuevo un importante crecimiento, producido por un incremento en el precio que compensa la pérdida en unidades en la mayoría de segmentos, en parte generado por el efecto de las medidas de desfinanciación de productos.

Si nos centramos ahora en el modelo de farmacia español, debemos decir que no ha estado exento de turbulencias en los últimos años. La primera década del siglo XXI ha supuesto para el sector farmacéutico de los países mediterráneos una época de intranquilidad pues su modelo de establecimiento farmacéutico ha sido cuestionado severamente por las autoridades de la Unión Europea.

Respecto a la planificación farmacéutica, una de las características distintivas del modelo español, la reglamentación de la Unión Europea no recoge ningún criterio comunitario, quedando a criterio del país correspondiente. Las directivas para la armonización de las legislaciones nacionales de los Estados miembros sobre el ejercicio profesional farmacéutico reconocen la competencia estatal en materia de establecimiento farmacéutico (72).

En este sentido, cada Estado somete a las oficinas de farmacia a una reglamentación propia atendiendo a sus peculiaridades poblacionales, geográficas y a su sistema sanitario. Esta circunstancia da lugar al mapa farmacéutico europeo que hemos analizado con anterioridad y que se especificaba en la Tabla 4, y del que se desprende que en Europa coexisten diferentes modelos de planificación de las oficinas de farmacia.

La Comisión Europea, ante legislaciones nacionales tan dispares y, sobre todo, con la clara intención de obligar a los Estados más restrictivos a aplicar un criterio liberalizador, en esta primera década del siglo XXI comenzó a cuestionar las reglamentaciones de determinados países y más concretamente los modelos de farmacia que establecen una regulación ordenada, ya que consideran que en ellos se establecen principios que vulneran el Derecho comunitario (13).

En particular, la Comisión Europea inició en junio de 2006 la tramitación de un procedimiento de infracción en aplicación del Artículo 226 del TCE5 contra España, Italia y Austria, por entender que algunos de los principios generales establecidos en sus respectivas legislaciones limitadoras del establecimiento de oficinas de farmacia, impedían garantizar el principio de libre establecimiento. Dichos principios son en concreto: reserva de la propiedad de farmacias para los farmacéuticos, exclusión de los farmacéuticos no nacionales de la propiedad de farmacias nuevas, preferencia por farmacéuticos con experiencia local, límites poblacionales para el establecimiento de nuevas farmacias y prohibición de la acumulación de propiedad de varias farmacias.

No obstante, el Tribunal de Justicia (Gran Sala) de la Unión Europea con fecha 1 de junio de 2010 (73), decretó que la reserva de la titularidad y la explotación de las farmacias a los farmacéuticos no vulnera el derecho de establecimiento, ni la libre circulación de

capitales. También declaró lícita la limitación en el número de autorizaciones para establecer farmacias en virtud de módulos poblacionales y distancias mínimas, e impide privilegiar en el proceso de selección a los farmacéuticos autóctonos. Dicho de otro modo, no es posible en la Unión Europea discriminar a los farmacéuticos por razón de nacionalidad, o por condiciones que tienen que ver con el desarrollo del ejercicio profesional en una determinada zona geográfica (74).

Aun así, la amenaza de un cambio en el modelo de farmacia mediterráneo siempre está presente en los foros y organizaciones colegiales (75) , (76).

Desde un punto de vista de la sociedad española, la farmacia goza de muy buena reputación. Así lo indican los estudios de percepción realizados sobre las farmacias.

En el año 2008 se realizó en Zaragoza un estudio de satisfacción del usuario de la oficina de farmacia (77), cuyos principales resultados respecto a la valoración sobre la farmacia fueron:

- Proximidad: el 85% de los consumidores compra en la farmacia que más cerca tiene de su domicilio.
- Fidelidad: el 77% de los consumidores compra siempre en la misma farmacia.
- Elevada satisfacción: el 99,9% de los usuarios está satisfecho con su farmacia habitual. La valoración media es de 8,1 (escala de 1 a 10).
- Trato y profesionalidad: los principales motivos de satisfacción (con la farmacia habitual) son el trato recibido y la profesionalidad o capacidad de resolución de dudas sanitarias.
- Sin necesidad de cambios: el 85% considera que no es necesario incorporar ningún cambio a las farmacias. El 8% cree que sí. Los principales cambios sugeridos son: disminuir precios, ampliar horarios y mejorar la disponibilidad de productos.
- Primer filtro sanitario: para dolencias leves, el 43% acude antes a la farmacia que al médico.

- Elevada disponibilidad: si falta un producto el 88% de los consumidores lo tienen el mismo día.
- Máxima confianza: para el 85% de los consumidores, las farmacias le proporcionan una elevada confianza.
- Horario adecuado: el 98% de los usuarios está satisfecho con el horario de las farmacias. La valoración media es de 7,5 (escala de 1 a 10).

Las encuestas sobre la valoración de la farmacia habitual realizadas por APROAFA en los años 2001 y 2009 (78) presentaron los resultados principales que aparecen en la Figura 42. En ella podemos observar que la valoración recibida por todos los ítems analizados es superior en el año 2009 con respecto al año 2001.

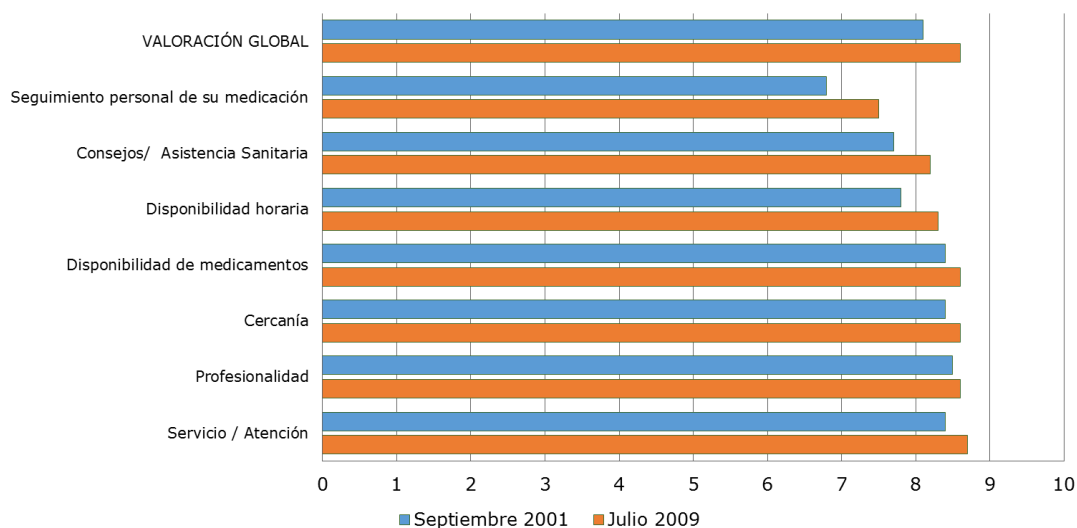


Figura 42. Valoración de la farmacia habitual en los años 2001 y 2009

El último estudio de percepción social de la oficina de farmacia realizado ha sido presentado en noviembre de 2014 por el Colegio Oficial de Farmacéuticos de Sevilla (81). Sus principales resultados son los siguientes:

- El 86% de la población declara tener una farmacia habitual o de cabecera, en tanto que sólo el 14% afirma ir a la farmacia que mejor le viene según el momento.

- El 30% de la población hace más de 20 años que conoce a su farmacéutico y este porcentaje sube al 43% en personas de más de 55 años. El estudio indaga en las razones de esta fidelidad de los usuarios a su farmacia y su farmacéutico, destacando la cercanía (en el 76% de las respuestas) y el trato (en el 42%) como principales razones.
- Establecimiento de referencia en el sistema de salud: el 89,9% la consideran muy o bastante importante para la salud pública, porcentaje que se incrementa hasta el 95,8% en las poblaciones pequeñas. Asimismo, en una de cada cinco encuestas, la farmacia sale como uno de los establecimientos que el encuestado visita por motivos de salud y en una de cada diez encuestas sale como la primera mención, por encima de médico de cabecera, consultas privadas y médicos especialistas, y sólo por debajo de centro de salud, ambulatorio y hospital. La frecuencia de visita es de 1,4 veces al mes, siendo las mujeres y los mayores de 55 años quienes más la visitan: 1,6 veces al mes. Los encuestados puntúan asimismo con un 8,45 sobre 10 la calidad del servicio prestado por las farmacias.
- Claves para elegir la farmacia: los factores de decisión principales son el componente humano, seguido del asesoramiento y conocimiento del farmacéutico y del personal que atiende.
- Impacto de la crisis: la encuesta muestra que los usuarios han valorado muy positivamente el esfuerzo de la oficina de farmacia por atenuar, eludir o contrarrestar el impacto de la crisis, tratando sobre todo de que no repercuta en la calidad de servicio.
- Modernización y receta electrónica: el estudio revela una clara percepción social de que las farmacias de Sevilla se han modernizado durante el periodo de crisis, lo que indudablemente redundará positivamente en la profesión. El 79% opina que lo han hecho además “mucho o bastante”. Uno de los factores que puede haber inducido a esa percepción de modernización es la dispensación mediante la receta electrónica, que se revela en la encuesta con un amplio índice de popularidad y respaldo. Así, más de un 75% de los usuarios conoce la receta electrónica y un 90,2% de quienes la conocen declaran utilizarla.

- El profesional farmacéutico: los usuarios destacaron, por este orden, la atención y trato al paciente; la formación, preparación y profesionalidad; el asesoramiento y consejo; y la amabilidad y simpatía.
- El consejo farmacéutico: casi el 75% de la población encuestada solicita consejo farmacéutico alguna vez, a menudo o muy a menudo, frente al 26,2% que lo pide rara vez o nunca, y que el 97% de los que suelen solicitar ese consejo lo valora muy positivamente. Los temas más recurrentes en los que se pide consejo farmacéutico son consultas sobre casos leves para evitar la visita al médico (en el 23,9% de las ocasiones); consultas para aclarar dudas sobre el tratamiento (en el 20% de los casos); y para informarse de la forma de administración de los medicamentos (en el 10% de los casos).

A modo de resumen de las cuatro encuestas analizadas sobre la percepción de la oficina de farmacia en España, podemos destacar los siguientes aspectos comunes:

- Valoración: la valoración general de los usuarios supera el 8 en todas las ocasiones.
- Proximidad: más del 85% de las personas encuestadas declara tener una farmacia de referencia que es la más cercana a su domicilio; la llamada “farmacia de cabecera”.
- Centro sanitario de referencia: la oficina de farmacia es identificada como un centro sanitario y en más del 10% de las ocasiones es visitado en primer lugar para resolver dolencias leves.
- El servicio y los profesionales: son altamente valorados principalmente por el trato recibido, la profesionalidad, la capacidad de resolución de dudas sanitarias, el asesoramiento y consejo y la disponibilidad de medicamentos.

Aunque no es un objetivo de esta investigación la realización de un diagnóstico de la situación de la farmacia, sí consideramos interesante utilizar la herramienta conocida como matriz DAFO (82). Esta herramienta presenta de un modo resumido los principales aspectos de un análisis de situación clasificándolos en: debilidades (D), amenazas

(A), fortalezas (F) y oportunidades (O).

A continuación, realizamos una definición de los cuatro elementos (81).

En el análisis de contexto o externo:

1. Oportunidades: son los factores externos positivos o favorables que pueden ser usados o aprovechados para el logro de la visión planteada. Son situaciones externas, es decir, que no son generadas por nosotros pero que pueden ser aprovechadas para el logro de algún objetivo.
2. Amenazas: son los factores externos negativos que limitan en el presente o pueden limitar en el futuro el logro de la visión de la organización. No son carencias que nosotros podamos tener, sino factores que no dependen de nosotros pero nos influyen.

En el análisis interno:

3. Fortalezas: Son las características positivas o habilidades que nosotros tenemos y que facilitan el logro de los objetivos propuestos. Pueden ser usadas tanto para aprovechar las oportunidades como para contrarrestar las amenazas.
4. Debilidades: Son las características negativas (carencias, limitaciones), que nosotros tenemos y que dificultan el cumplimiento de nuestros objetivos. Son factores que nos pueden hacer perder las oportunidades que se nos presentan y que nos hacen vulnerables ante las amenazas.

Como hemos visto, las fortalezas y debilidades se corresponden con el ámbito interno y las amenazas y oportunidades con el ámbito externo.

Como resumen de la revisión bibliográfica realizada en esta investigación y de todo lo expuesto hasta el momento, podemos confeccionar la matriz DAFO de la oficina de farmacia española que observamos en la Figura 43.

| Debilidades | Amenazas |
|---|--|
| <ul style="list-style-type: none"> • Cultura empresarial mejorable por parte del farmacéutico • Bajo uso de herramientas de gestión empresarial • Insuficiente aplicación de las nuevas tecnologías • Alta dependencia del mercado público regulado • Venta libre inferior a lo deseable • Baja diferenciación entre farmacias | <ul style="list-style-type: none"> • Liberalización del sector • Reducción del gasto público • Medidas de contención de la demanda de medicamentos • Retrasos en los pagos • Salida de medicamentos de la oficina de farmacia hacia la dispensación hospitalaria • Aumento de franquiciados en parafarmacia • Disminución del precio del medicamento financiado • Aparición de cadenas de farmacia |
| Fortalezas | Oportunidades |
| <ul style="list-style-type: none"> • Prestigio de la profesión • Alta valoración de los usuarios • Cercanía y proximidad • Consejo y trato personalizado • Referencia para los pacientes en su salud • La oficina de farmacia es el establecimiento sanitario más cercano al ciudadano • Imagen de la farmacia en la población | <ul style="list-style-type: none"> • La población está cada vez más comprometida con el autocuidado de la salud y la prevención • Especialización de las oficinas de farmacia • Aplicación de herramientas de gestión empresarial y marketing • Aplicación de las nuevas tecnologías en el conocimiento de los clientes • Aplicación de técnicas de segmentación y fidelización de clientes • Potencial desarrollo de nuevos servicios remunerados • Mayor aprovechamiento de la receta electrónica • Medicina personalizada • La comunicación médico-farmacéutico • Venta de medicamentos sin receta por internet |

Figura 43. Matriz DAFO de la oficina de farmacia española

1.2.7 Retos y tendencias de futuro analizados

Según los informes de *IMS Health* (82) , (65) sobre tendencias del mercado farmacéutico español, los factores de contribución positivos o negativos del mismo son los que se aprecian en la Figura 44.

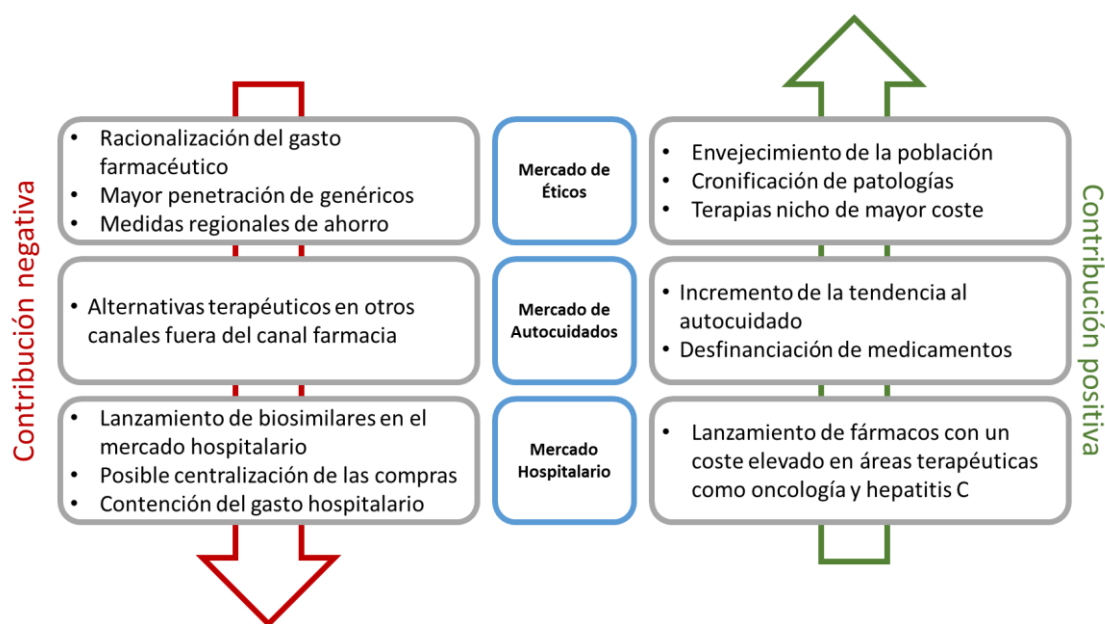


Figura 44. Factores que incidirán en el mercado farmacéutico español de forma positiva o negativa

Estas tendencias junto con la información de la Figura 43, se tendrán en cuenta en esta investigación a la hora de realizar la propuesta sobre acciones y herramientas encaminadas a salvar las dificultades de las farmacias y a potenciar los factores positivos presentes en los mercados donde opera la oficina de farmacia española.

La información expuesta con anterioridad en esta investigación, pone de manifiesto que el sector farmacéutico está experimentando un cambio muy evidente que debe ser afrontado aprovechando las oportunidades que este cambio está provocando, aprovechando las fortalezas, huyendo de las amenazas en la medida que sea posible y minimizando las debilidades.

En este contexto, es lógica la perturbación que sufren los farmacéuticos porque van a tener que adaptarse a múltiples cambios, a nuevos competidores y a nuevas formas de gestión y relación con sus clientes. Por una parte, se tienen que enfrentar a la

transformación de su sector a lo que hay que sumar las nuevas reglas del mercado, con un consumidor más exigente, más infiel, con menor capacidad de gasto y más informado, y la revolución de las nuevas tecnologías. Ya no sirve seguir haciendo lo mismo. A partir de ahora hay que contar con una verdadera estrategia de marketing para afrontar el reto de dar el salto de farmacéuticos a empresarios. Algunos ya han comenzado a desarrollar estrategias de marketing adaptadas a su empresa, término éste que, en muchos casos, por el tipo de servicio que prestan, son reticentes a utilizar aplicado a su farmacia y que, en absoluto, debe estar reñido con la labor sanitaria que llevan a cabo (83).

1.3 Herramientas para el desarrollo de la oficina de farmacia

Llegados a este punto, y una vez analizada la situación actual y los posibles escenarios futuros a los que se enfrenta la oficina de farmacia, resulta obvio esta debe salir de la llamada zona de confort para visitar su zona de aprendizaje y desarrollo.

En este apartado, trataremos a la oficina de farmacia como una empresa. Teniendo en cuenta las siguientes apreciaciones (84):

- a) La oficina de farmacia es una empresa privada. El servicio sanitario que la oficina de farmacia presta no se concibe ni es realizable sin un simultáneo ejercicio comercial de carácter instrumental.
- b) La naturaleza del servicio sanitario no desvirtúa el carácter mercantil de la actividad que se lleva a cabo en la oficina de farmacia que, en cambio, simplemente predetermina su funcionamiento, debilitándose e incluso eliminándose, el aspecto comercial de la actividad en la medida que constituya obstáculo o desviación con relación al aspecto sanitario de la propia actividad o lo contradiga.

Según el régimen establecido en la normativa básica, la actividad farmacéutica es una actividad sanitaria sujeta al doble principio de ejercicio libre de la profesión y de libertad de empresa, pero ello no significa que el profesional titulado que la preste sea un funcionario público. Al contrario, es un profesional libre que ejerce una profesión libre como las demás que también lo son. Además, dicha actividad puede y tiene que ser sometida a determinados requisitos y limitaciones para salvaguardar el interés público concretado en la protección de la salud según el artículo 43 de la Constitución Española (85).

En los apartados anteriores hemos podido ver cómo está conformado el mercado farmacéutico en el que realiza su actividad la oficina de farmacia, su evolución y sus agentes.

Desde el punto de vista de una empresa, la oficina de farmacia opera en dos mercados bien distintos entre sí:

- El mercado de medicamentos de prescripción (RX).
- El mercado del autocuidado de la salud (OTC y productos no medicamentos).

El primero de ellos, es un mercado regulado y sin posibilidad de aplicar herramienta alguna salvo aquellas propuestas por la Administración como lo es la Receta Electrónica. Por lo tanto, todo lo que se exponga a partir de este momento en la investigación solo tendrá aplicación en el mercado del autocuidado de la salud y de los servicios.

En el mercado del autocuidado de la salud, el farmacéutico sí encuentra una competencia clara. Aunque la oficina de farmacia parte con ventaja gracias a su imagen y a operar en solitario en el mercado de los medicamentos de prescripción, del que provienen la mayoría de medicamentos presentes en el mercado del autocuidado, se enfrenta a empresas de ámbito exclusivamente privado acostumbradas a competir y a poner en marcha prácticas empresariales comerciales y de marketing en la que son expertos frente a la inexperiencia de las oficinas de farmacia en estas lides.

Cuando se pregunta a un empresario acerca de los principales escollos o dificultades que atentan contra el crecimiento de su empresa (86), las respuestas que se obtienen son muy variadas pero lo general suelen rondar temas comunes: la falta de seguridad jurídica, la imprevisibilidad de las decisiones gubernamentales, el alto costo fiscal, la dificultad de contar con personal capacitado, la ausencia de financiamiento adecuado, que no barato, etc. Lo que posiblemente nadie señale como un freno al desarrollo es lo que se conoce como “zona de confort”. Todas las personas tendemos a ubicarnos en nuestra zona de confort, ese espacio no necesariamente físico, en el cual nos sentimos cómodos, seguros y confortables. Es lo que sabemos hacer, y más aún, lo que nos gusta hacer.

Para contrarrestar los factores negativos del mercado que hemos visto con anterioridad, es necesario que el farmacéutico haga una incursión en la zona de aprendizaje desde su zona de confort (Figura 45). La idea es que trabajando en la zona de aprendizaje se amplíe su zona de confort (87).

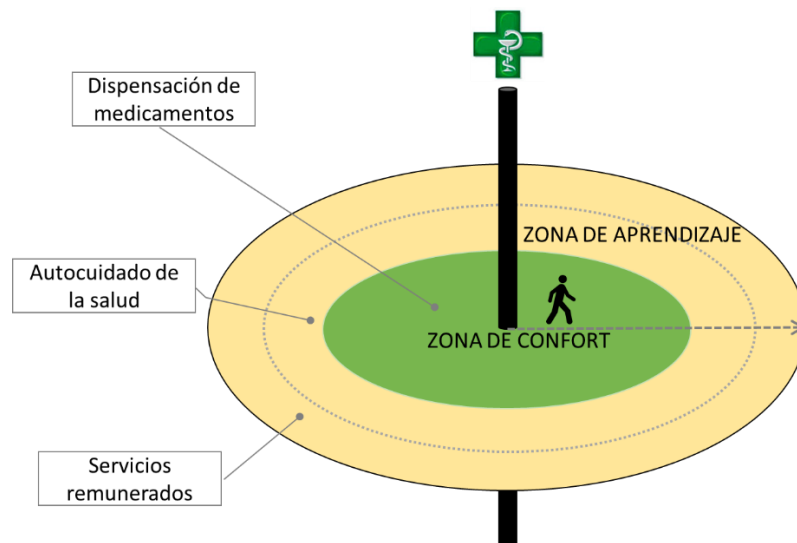


Figura 45. Zona de confort y zona de aprendizaje

En los siguientes párrafos trataremos varias herramientas para el desarrollo de la oficina de farmacia y soporte al viaje a la zona de aprendizaje.

1.3.1 Productos y servicios de la oficina de farmacia

En este apartado, nos centraremos en la oficina de farmacia como establecimiento, con la idea de conocer su funcionamiento y los productos y servicios ofrecidos.

En la Figura 46 podemos observar una clasificación de los productos que podemos encontrar en una oficina de farmacia (88):

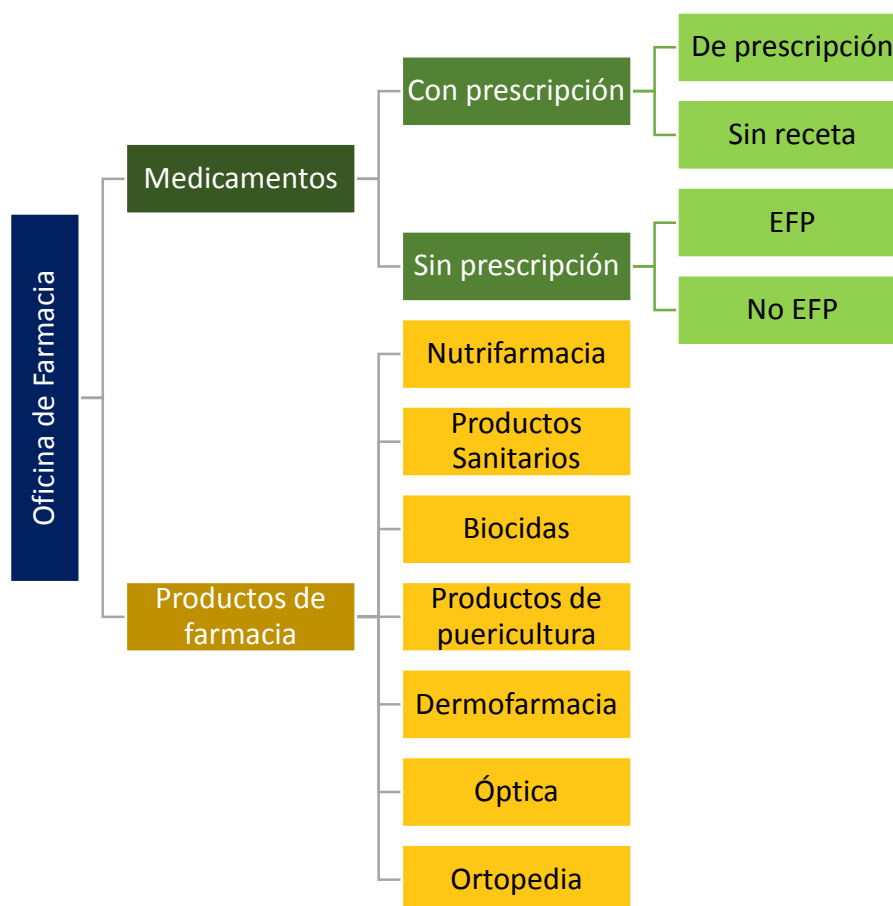


Figura 46. Clasificación de los productos en una oficina de farmacia

Los productos que se han identificado como de parafarmacia en la Figura 46, también pueden ser adquiridos en una parafarmacia o establecimiento similar de los que aparecen en la Figura 47 (excepto los dietoterápicos, que son complementos dietéticos destinados a personas que padecen algún problema de salud).

De hecho, estos productos son aquellos en los que las oficinas de farmacia encuentran a una competencia más preparada desde el punto de vista comercial y de marketing. Y, como hemos visto anteriormente, puede ser un área de expansión posible del negocio de las oficinas de farmacia.

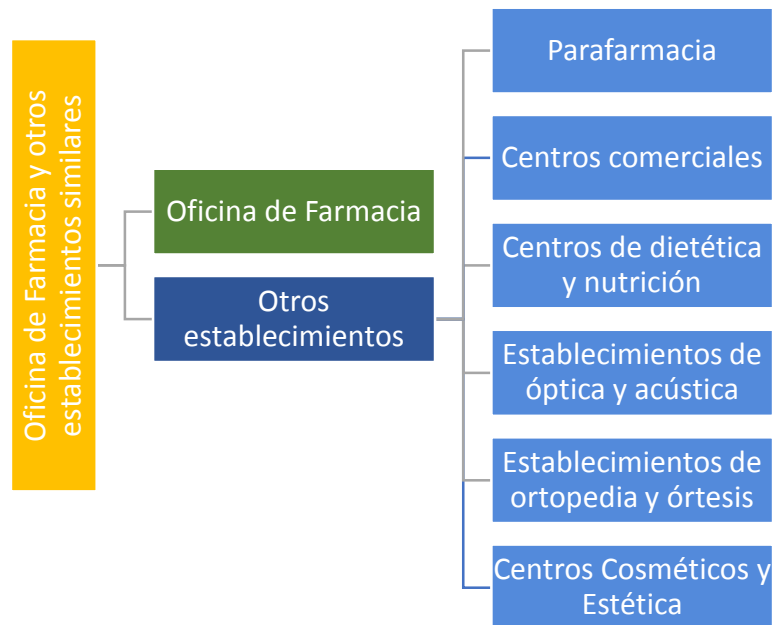


Figura 47. Oficina de Farmacia y otros establecimientos similares

Respecto a los servicios que se prestan en la actualidad en las oficinas de farmacia y según el Consejo General de Colegios Oficiales de Farmacéuticos (91), en base a su naturaleza, atienden a la siguiente clasificación:

| CLASIFICACIÓN DE LOS SERVICIOS EN UNA OFICINA DE FARMACIA | | | | |
|---|--|---|--|---------------------------------|
| Clasificación | Descripción | Naturaleza | Obligatoriedad | Requerimientos |
| Servicios esenciales o básicos | Son los servicios inherentes a la condición de farmacéutico de farmacia y que están definidos como sus funciones básicas | Desarrollo de funciones básicas de la farmacia | Si, para la totalidad de las farmacias | Sin requerimientos adicionales |
| Servicios complementarios | Son aquellos servicios relacionados con programas específicos de salud pública que el farmacéutico está legitimado a prestar, siempre que se cumplan los requisitos establecidos por las autoridades competentes | Participación en programas específicos de Salud Pública | Son voluntarios | Previa formación y acreditación |
| Servicios colaborativos | Son aquellos servicios que se desarrollan en un ámbito local en colaboración con otros profesionales sanitarios de su entorno y/o con el paciente. Se relacionan con el paciente crónico y sus patologías | Colaboración entre profesionales | Son voluntarios | Previo convenio y acreditación |

Tabla 34. Clasificación de los servicios en una oficina de farmacia

Asimismo, en la Figura 48 se muestran los 15 servicios que atienden a la clasificación de la Tabla 34.

| Clasificación | | Concepto | Servicio |
|--------------------------------------|--------------------------------|---|---|
| Servicios en una oficina de farmacia | Servicios esenciales o básicos | Custodia de medicamentos | Adquisición y control de medicamentos |
| | | Elaboración de medicamentos | Formulación magistral |
| | | Dispensación de medicamentos | Dispensación de medicamentos y productos sanitarios |
| | | Indicación farmacéutica | Indicación farmacéutica |
| | | Información personalizada del medicamento | Información y verificación del cumplimiento del tratamiento |
| | | Farmacovigilancia | Detección y notificación de efectos adversos |
| | | Soporte al autocuidado | Educación en ámbito de la salud pública |
| | Servicios complementarios | Programas específicos de salud pública | Identificación de factores de riesgo cardiovascular |
| | | | Identificación precoz de patologías |
| | | | Programa de mantenimiento con metadona |
| | | | Apoyo y formación a escuelas y entidades asociativas |
| | Servicios colaborativos | Uso de medicamentos | SPD |
| | | | Soporte en el tratamiento de inicio |
| | | | Seguimiento farmaterapéutico |
| | | Autocontrol | Verificación de parámetros biológicos |

Figura 48. Servicios ofrecidos en una oficina de farmacia

Si por cartera de servicios tomamos la definición del observatorio de la cartera de servicios (90): “conjunto de servicios farmacéuticos ofrecido desde la Oficina de Farmacia, paralelo a la dispensación de medicamentos y fundamentado en un protocolo que prevé la acreditación y la formación del farmacéutico, y que tiene por objetivo mejorar la salud de la ciudadanía, atendiendo a sus necesidades reales”, la Figura 48 refleja la cartera de servicios de la oficina de farmacia española.

1.3.2 Pacientes y clientes: usuarios

A lo largo de la revisión de la bibliografía e información existente sobre la evolución de las oficinas de farmacia desde su creación, podemos observar que el concepto paciente resulta un denominador común. Sin embargo, en los últimos tiempos está apareciendo el concepto de cliente como forma de identificar a aquellas personas que visitan la farmacia a comprar productos y servicios no necesariamente del tipo medicamento. La creación de esta nueva categoría permite la inclusión de los métodos y herramientas del llamado marketing farmacéutico, abriéndose un mundo lleno de posibilidades ante los farmacéuticos para evolucionar en la oficina de farmacia una línea tradicionalmente alejada para y por el propio colectivo: la venta libre.

Una vez aceptado que la oficina de farmacia está experimentando una aceleración en su proceso evolutivo parece oportuno disponer de una clara identificación de los usuarios de los productos y servicios de las oficinas de farmacia.

En este apartado vamos a distinguir adecuadamente los términos cliente y paciente dado que aunque similares en el uso normal del lenguaje, presenta diferencias notables en el entorno socio-sanitario y concretamente en el sector farmacéutico.

Dado que hay cierta resistencia a la hora de considerar el término “cliente” en el ámbito farmacéutico, vamos a separar ambos conceptos para que quede claro que las herramientas para el desarrollo de farmacia que tratamos en esta investigación tienen como destinatarios a los clientes y en ningún caso a los pacientes. Aunque, como veremos más adelante, puede ocurrir que un cliente sea a la vez paciente, aun así, nuestro objetivo siempre será el cliente.

Del latín *patiens* (“padecer”, “sufrir”), paciente es un sustantivo que hace referencia a quien tiene paciencia, a la capacidad de soportar o padecer algo, de hacer cosas minuciosas o de saber esperar. Según la RAE (93), se trata de una persona que padece física y corporalmente, y especialmente quien se halla bajo atención médica.

En nuestro caso podemos decir que paciente es la persona que recibe los servicios de un médico u otro profesional de la salud y se somete a un examen, a un tratamiento o

a una intervención y que puede acudir a una oficina de farmacia a retirar sus medicamentos.

Respecto al término cliente, la *American Marketing Association* (A.M.A.) (92) lo define como "el comprador potencial o real de los productos o servicios". Según *The Chartered Institute of Marketing* (93), el cliente es "una persona o empresa que adquiere bienes o servicios y no necesariamente como el consumidor final"

Otra definición en el ámbito del marketing que encontramos en la literatura (94) es la siguiente: "término que define a la persona u organización que realiza una compra. Puede estar comprando en su nombre, y disfrutar personalmente del bien adquirido, o comprar para otro".

Como vemos, el término "cliente" permite hacer mención a la persona que accede a un producto o servicio a partir de un pago, sea para él o no. A veces, y según el contexto, puede ser utilizado como sinónimo de comprador (la persona que compra el producto), usuario (la persona que usa el servicio) o consumidor (quien consume un producto o servicio).

A veces se presenta cierta controversia en el uso del término "paciente" en el entorno farmacéutico, ya que algunos consideran que "paciente" es un término exclusivo del entorno médico y por ello se utiliza la palabra "cliente". En nuestro caso, dejaremos clara la diferencia entre ellos en lo que respecta al entorno de la oficina de farmacia.

Según las prestaciones que se ofrecen en las oficinas de farmacia, asociaremos el concepto de "paciente", a la persona que tiene un diagnóstico hecho y un tratamiento instaurado y el farmacéutico ejerce parte de la atención: dispensación segura, soporte en el conocimiento y buen uso del medicamento, soporte en la farmacovigilancia y ayuda en la mejora de la adherencia. Por "cliente" entenderemos el sujeto que recibe un producto o servicio relacionado con el bienestar que no requiere diagnóstico previo ni tratamiento concreto y que paga un precio por ello (95).

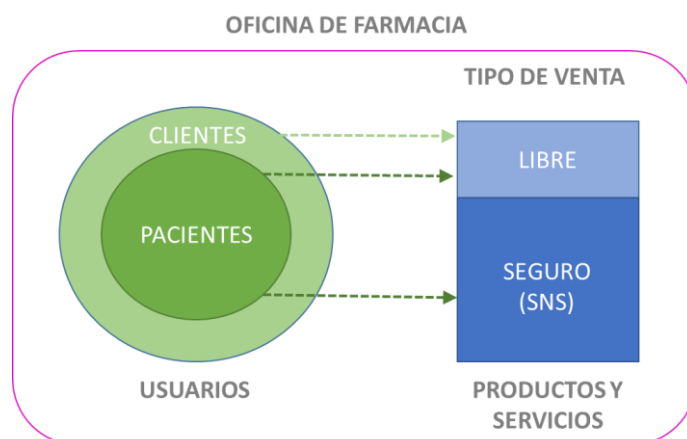


Figura 49. Tipos de usuarios en una oficina de farmacia

Según esta clasificación, y usando teoría de conjuntos, podemos afirmar que el conjunto de pacientes es un subconjunto del conjunto de clientes como se aprecia en la Figura 49. En la actualidad, en las oficinas de farmacia entran muchos más pacientes que clientes, esto explica el mix de ventas (70% dispensación de medicamentos y 30% venta libre), hecho que justifica la utilización de nuevas herramientas por parte del farmacéutico que balancee las ventas hacia la venta libre sin olvidar que el “paciente” es el eje principal del desarrollo profesional y asistencial del farmacéutico.

Ahora bien, independientemente que se trate de un cliente o un paciente, un hecho de vital importancia para la oficina de farmacia es el registro de la actividad de cada uno de ellos. En el caso de los pacientes, es más sencillo puesto que iniciativas como la receta electrónica proveen de información al farmacéutico. En el caso de los clientes se dispone de menos información registrada aunque en los últimos tiempos, gracias al uso de las tarjetas de fidelización, se va disponiendo de más registros sobre la actividad de los clientes en las oficinas de farmacia.

Un resumen de la clasificación de pacientes-clientes, el tipo de venta y si la información se registra o no, podemos verlo en la Figura 50.

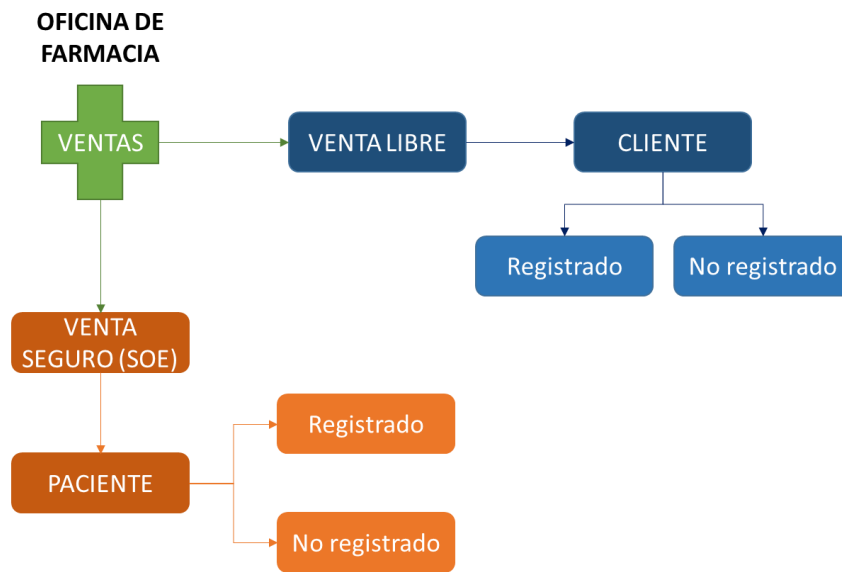


Figura 50. Tipología de ventas y usuarios en una oficina de farmacia

Por último, y no menos importante, desde esta investigación no queremos pasar por alto la importancia que tiene la identificación unívoca de pacientes y clientes. En el caso de los pacientes es más común que este problema esté bien resuelto ya que normalmente, al menos en España, el paciente está identificado por su número de historia de salud. Aun así, es importante destacar, en algún sistema sanitario puede ser hoy todavía un problema, que el identificador debe ser correspondiente con el número de historia, que será la clave principal de la entidad, y, por tanto, no debe depender de los valores del atributo del elemento que identifica y preferiblemente que sea un número secuencial “no parlante”. El debate acerca de si debe utilizarse un identificador universal (DNI, número seguridad social, carnet de conducir, etc.) o un identificador específico para la atención sanitaria (número de historia) es resuelto de manera diferente en cada país europeo: Francia, Alemania, Grecia y España utilizan identificadores específicos, al contrario que los países escandinavos y del oriente europeo.

1.3.3 Marketing farmacéutico

Como ya avanzamos en capítulos anteriores, la oficina de farmacia debe entenderse bajo el prisma de una empresa que está inmersa en un entorno competitivo concreto y de cambio, en el que es necesario encontrar un posicionamiento empresarial diferenciado. La farmacia debe, por tanto, conocer en qué mercado y, por extensión, cuál es el entorno en el que desarrolla su actividad. Este análisis es previo a consideraciones de un nivel superior, como el estudio de la demanda o su incidencia a través de los diferentes instrumentos del marketing (producto, precio, distribución y promoción).

El mercado de la farmacia se caracteriza por la existencia de un gran número de oficinas que pueden desarrollar o desarrollan actuaciones muy similares. Por tanto, en este marco se hará difícil mantener posiciones de liderazgo y el éxito de las actuaciones de las oficinas de farmacia vendrá establecido principalmente por la diferenciación de los productos y servicios que se ofrezcan y por el grado de fidelización conseguido con sus clientes (96).

Entre las muchas definiciones de marketing existente tomamos, por ejemplo, la facilitada por la Asociación Americana de Marketing (AMA, por sus siglas en inglés) que tiene una gran ascendencia en esta disciplina y cuya definición de marketing es: “la actividad, conjunto de instituciones y procesos para crear, comunicar, entregar, e intercambiar ofertas que tienen valor para los clientes, socios y la sociedad en general” (92). Es común confundir marketing y ventas, puesto que las ventas son las manifestaciones tangibles del marketing; sin embargo, no son su punto focal, el marketing lo que busca en primera instancia es identificar las necesidades del consumidor, para poder crear un producto que le pueda satisfacer a un precio aceptable y explorando formas de promocionar el producto de una manera eficaz.

El marketing tiene como objetivo principal identificar las necesidades de los consumidores y desarrollar productos en base a ello. Así mismo, debe fijar un precio justo y adecuado, sin dejar atrás lo que es la promoción y distribución del mismo para lograr las ventas deseadas.

El paradigma tradicional del marketing se caracteriza por una orientación de tipo transaccional. La idea central es que se circunscribe a la mezcla de marketing, a través de la mixtura de cuatro variables (producto, precio, promoción y distribución) que convence al consumidor para que compre un determinado producto (99). En contraposición al paradigma tradicional transaccional, una nueva aproximación al marketing, basada en las relaciones, ha ido tomando fuerza. La globalización, la intensidad competitiva, los cambios del consumidor, el perfeccionamiento de las estrategias del marketing, los avances tecnológicos, la preocupación por la calidad y los cambios adicionales producidos en el entorno en el que operan las organizaciones han llevado a que el enfoque transaccional del marketing no solucione los problemas actuales y haya sido sustituido por una perspectiva relacional (98). En el actual entorno competitivo, el cliente se ha convertido en el elemento más escaso del sistema, siendo su conservación, y no su captación, la clave del éxito empresarial.

El Marketing Relacional parte de la premisa de que toda actividad comercial es como la vida misma: se basa en relaciones, y para ello utiliza todas las técnicas a su alcance, incluyendo las últimas tecnologías en comunicación y producción para convertir cualquier contacto con un cliente real o potencial en una relación duradera y satisfactoria para la marca y el consumidor (99).

A muchos farmacéuticos conceptos como el “plan de marketing”, el “plan financiero”, o el “plan de comunicación”, etc. les siguen pareciendo lejanos o completamente ajenos, porque los asocian al mundo de las grandes corporaciones multinacionales. El marketing, en concreto, ha tenido una historia reciente de recelos y desconfianza por parte de la profesión farmacéutica. Históricamente se ha considerado que no era ético promover la “venta” de un servicio farmacéutico, olvidando que en realidad, la atención a la salud se desarrolla en un ámbito económico y de mercado.

Las técnicas de marketing pretenden justamente dar a conocer el valor diferencial de nuestros productos y servicios a nuestros clientes potenciales, y en consecuencia, el prestigio y la profesionalidad de los farmacéuticos y personal de la oficina de farmacia que les atenderán. Conocer los puntos fuertes y débiles de la actividad de la oficina de

farmacia, analizar la competencia (fuera del ámbito de los medicamentos) y las necesidades de los clientes, construir un posicionamiento, estructurar el servicio, velar por la atención a los clientes, comunicarse con ellos, etc., son tareas estratégicas que también debería desarrollar una oficina de farmacia.

La creciente exigencia del mercado y la transformación tecnológica están generando mayor conciencia de la necesidad del marketing como herramienta para identificar y conectar con nuestros clientes potenciales para ofrecerles soluciones a sus problemas de salud de una forma participativa, transparente y divulgativa.

El marketing en el sector farmacéutico está marcado por un factor fundamental que requiere una metodología muy específica y diferente de la que se aplica en la mayoría de sectores no sanitarios. En el ámbito farmacéutico, el cliente es alguien que no se sabe cliente y que sólo se lo plantea cuando tiene una necesidad que normalmente está relacionada con su salud. En este sentido, todo el empeño en el marketing farmacéutico debe ser conseguir que la farmacia sea una opción para el cliente en el momento que se le presente la necesidad sobre el autocuidado de su salud.

El marketing no son sólo campañas de publicidad, promociones o descuentos orientados a conseguir resultados en el corto plazo. El marketing debe entenderse como un sistema estratégico integrado que permita ordenar y priorizar en el tiempo las herramientas y acciones necesarias para dar a conocer nuestros productos y servicios de valor a nuestros clientes.

Con cuidado exquisito de no traspasar la línea que separa los términos “clientes” y “pacientes”, se trata de crear y aplicar una metodología contrastada en otros sectores y basada en métricas que nos permitan tomar decisiones activando mecanismos que impliquen a todos los miembros de la oficina de farmacia.

Dado que el farmacéutico titular es a la vez propietario (empresario) de la oficina de farmacia, resulta de vital importancia identificar su perfil previamente a la implantación de estas nuevas herramientas. Según estudios realizados en el sector (100), los perfiles actitudinales de los farmacéuticos se clasifican en cuatro grupos:

- Farmacéutico tradicional: es aquel depositario del significado histórico de la farmacia. Se corresponde por lo general, con farmacéuticos mayores, con un alto grado de compromiso y con un fuerte componente de resistencia y lucha.
- Farmacéutico dispensador: de perfil pasivo, no suele realizar recomendaciones a no ser que le pidan opinión. Según los estudios, persigue especialmente el bienestar y la comodidad. Son farmacéuticos mayores que no quieren conflicto ni compromiso.
- Farmacéutico sanitario: de perfil “consejero sanitario”, centra su negocio en la atención sanitaria al paciente y se ha adaptado a la tecnología. Son, tanto farmacéuticos jóvenes como mayores, con capacidad de adaptación al cambio.
- Farmacéutico empresario: este perfil surge fundamentalmente de la nueva generación en el que el farmacéutico desempeña un rol de gestor y empresario. Su visión del negocio se centra en los márgenes comerciales y en ofrecer servicios al cliente, no sólo sanitarios. Es el que más ha apostado por el uso de las nuevas tecnologías, por la ampliación de horarios, por la especialización y por la diferenciación.

Es evidente que a la vista de la revisión de la situación actual y del tipo de perfil de un farmacéutico, las herramientas propuestas en esta investigación tendrán una mayor o menor acogida. Lo que queda fuera de toda duda es que la farmacia está en plena evolución y, por lo tanto, debe rodearse de nuevas técnicas y herramientas que la haga más fuerte para afrontar este proceso con mayor garantía.

1.3.3.1 Fidelización de clientes

Actualmente, las empresas intentan desarrollar lazos más fuertes de lealtad con sus clientes finales. En el pasado, muchas empresas consideraban que el número de sus clientes estaba garantizado. Los clientes no tenían muchas alternativas, o los otros competidores no eran buenos en calidad o servicio, o el mercado crecía tan rápidamente que la empresa no se preocupaba por conseguir una alta satisfacción de sus clientes

puesto que los iba recuperando al ritmo que los perdía. Esta situación parte de la teoría de que siempre que se va un cliente se puede reemplazar por otro nuevo. Es evidente, que este supuesto ya no es tan trivial como antes. Hoy en día, cada vez es más ardua la tarea de complacer a los clientes. Tienen más información, son más conscientes de las diferencias de precios, más exigentes, menos tolerantes y están abordados por más competidores que les hacen ofertas mejores o similares. El desafío, según Jeffery Gitomer (101), no consiste en lograr clientes satisfechos; muchos competidores pueden hacer lo mismo. El desafío consiste en lograr clientes leales.

Los estudios realizados en los últimos años han enfatizado los claros beneficios de fidelizar a nuestros clientes frente a la captación de clientes nuevos. Uno de esos estudios, realizado por la oficina americana del consumidor, concluye que cuesta cinco veces más conseguir un nuevo cliente que mantener un cliente actual (104). Según Luis Huete (103) la fidelización de clientes presenta dos dimensiones: una subjetiva y otra objetiva. La subjetiva se centra en establecer vínculos emocionales entre el cliente y la empresa. Por el contrario, la objetiva está relacionada con el perfil del comportamiento del cliente, con lo observable, lo medible, esto es, lo objetivable.

En esta investigación nos centraremos en su dimensión objetiva. Si bien es cierto que ésta se apoya en la subjetiva, esto podrá tratarse en futuras investigaciones.

Una empresa fideliza a sus clientes en la medida que es capaz de lograr una alta tasa de retención. Se define el concepto retención como la repetición de una compra, o de un volumen mínimo prefijado, durante un determinado periodo de tiempo. Fidelizar es obtener tasas altas de clientes que repiten o tasas bajas de no repetidores o desertores. Por otro lado, la tasa de deserción es el porcentaje de clientes que perdemos entre un periodo y otro. La relación entre ambas tasas se muestra en la ecuación (1):

$$\text{Tasa de retención (TR)} = 1 - \text{Tasa de deserción} \quad (1)$$

Es importante observar que la tasa de deserción no mide la diferencia entre la cartera de clientes entre un periodo y otro. Asimismo, resulta fundamental definir con exactitud qué significa el término deserción en el seno de la empresa. El abandono de una

empresa por parte del cliente no tiene por qué limitarse a la marcha a la competencia; es algo más sutil. Por ejemplo, si un banco midiese el abandono de sus clientes como el número de cuentas corrientes que se cierran, estaría falseando inconscientemente su tasa de deserción. Es evidente, que la frecuencia o volumen de compra son aspectos a tener en cuenta para definir una deserción. Un segundo aspecto a tener en cuenta es determinar el periodo que debe transcurrir para considerar que un cliente nos ha abandonado. Como en el primer aspecto, la elección del periodo depende del tipo de empresa; no será lo mismo en una oficina de farmacia que en un hotel invernal. Una de las formas de establecerlo consiste en agregar a la frecuencia de compra de los clientes habituales un tiempo extra como margen de seguridad. Este tiempo extra dependerá del perfil de la empresa y de lo rigurosa que quiera ser en este sentido.

Una expresión útil para obtener la tasa de deserción aparece en la ecuación (2):

$$\text{Tasa de deserción} = 1 - \frac{\text{Clientes } (t \rightarrow t+1)}{\text{Clientes } (t)} \quad (2)$$

Un gran reto por parte de las empresas surge a la hora de afrontar una disminución en su tasa de deserción. Su impacto en la rentabilidad está bien estudiado empíricamente. Un ejemplo del mismo puede encontrarse en la investigación realizada por F. Reichheld y W. Sasser (104) en Estados Unidos, que concluye que con el mismo número de clientes totales, la disminución de un 5% en la tasa de deserción produce un impacto en los beneficios de la empresa que oscila, según los sectores, entre un 30% y un 50%, como mínimo.

Por último, la tasa de deserción también nos permite calcular un nuevo indicador: la vida media de un cliente:

$$\text{Vida media de un cliente} = \frac{1}{\text{Tasa de deserción}} \quad (3)$$

, donde la tasa de deserción está expresada en porcentaje.

Así por ejemplo, si la tasa de deserción media de los clientes de una empresa es del 10%, esto equivale a una vida media del cliente de 10 años. Lo que supone que en 10 años la

empresa debe renovar completamente su cartera de clientes. Si la unidad de tiempo determinada por la empresa no fuera el año, el cálculo se haría de la misma forma solo que con la unidad correspondiente y la tasa de deserción haciendo referencia a dicha unidad temporal. Podemos ver el comportamiento de este indicador en la Figura 51.

El indicador “vida media de un cliente” permite una mejor visualización de la importancia de la fidelización. La razón es sencilla: si fidelizar a nuestros clientes es conseguir que nuestros clientes vuelvan de forma continuada, debemos disminuir la tasa de deserción ya que esto conducirá a un aumento de la vida media de los clientes. Por lo tanto, podemos concluir que fidelizar es alargar la vida media de los clientes y reducir el número de desertores.

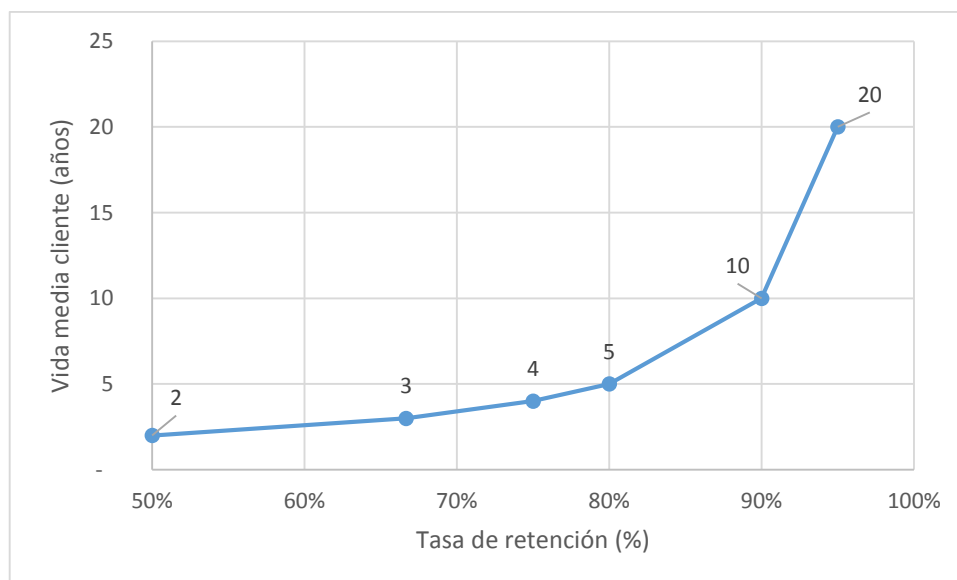


Figura 51. Relación entre la vida media y la tasa de retención de clientes

Además de la vida media de un cliente como indicador operativo de fidelización, existen otros indicadores que se recogen en la Tabla 35. La elección de uno u otro depende de los factores que la empresa quiera mejorar.

| Indicador | Descripción |
|-------------|---|
| Retención | Porcentaje de clientes que son fieles porque están satisfechos (retenidos) |
| Novedad | Tiempo desde la última venta/contacto |
| Frecuencia | Tiempo entre dos ventas/contactos |
| Antigüedad | Fecha de inicio |
| Cantidad | Pedido medio por periodo |
| Referencias | Porcentaje de clientes que vienen por referencias de otros clientes/amigos y familiares |

Tabla 35. Otros indicadores de fidelización

En la aplicación de las técnicas del marketing relacional es fundamental saber qué es lo que valora el cliente en su relación con nuestra empresa. El valor es un aspecto subjetivo que depende de cada uno de nosotros y dado que nuestros intereses, motivaciones y estados de ánimo son temporales, el valor no deja de ser un aspecto dinámico.

Entre las diferentes definiciones de valor para el cliente elegimos la que aparece en la Figura 52 (105).

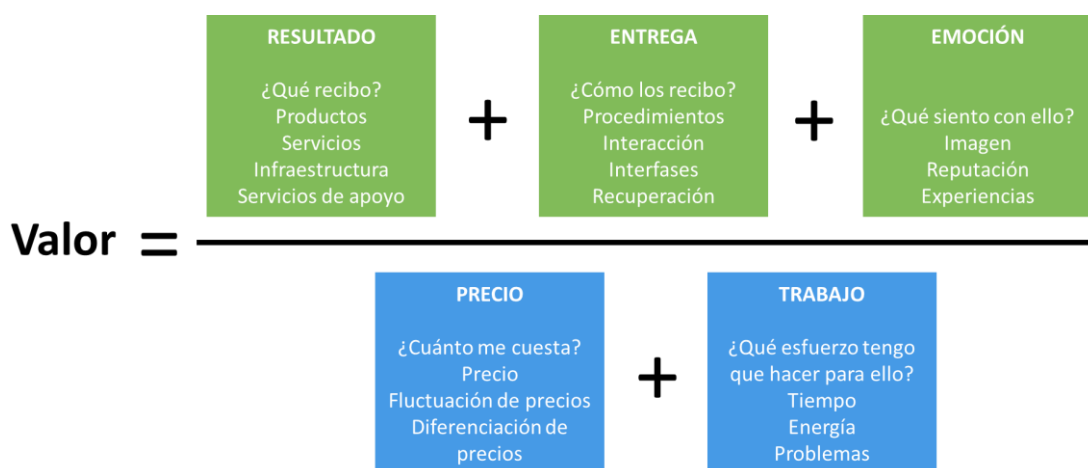


Figura 52. Ecuación de valor para el cliente

En ella podemos ver que el valor resulta del cociente entre lo que obtiene el cliente (resultado, entrega y emoción) y lo que entrega a cambio (precio y trabajo)

Llegados a este punto, ya estamos en disposición de hablar sobre el valor del ciclo de vida del cliente (*Customer Lifetime Value*, CLV) que está ganando cada vez mayor importancia como indicador de marketing. El *Customer Lifetime Value* se define como el

valor actual de los futuros beneficios aportados por el cliente durante su relación con la empresa. Hablando de forma sencilla, el *CLV* es el valor neto de los ingresos que un cliente nos genera durante el tiempo en que el cliente es cliente nuestro. Se trata, por tanto, de valorar a cada cliente en función de su aportación al margen neto con sus compras, la frecuencia con la que realiza dichas compras, la fidelidad prevista representada por la tasa de retención, y los costes generados en la gestión de la relación con el consumidor. Va más allá de la aplicación de un modelo matemático que reproduzca la realidad de una forma aproximada y resulta muy interesante la información proporcionada por este indicador a la hora de la toma de decisiones.

Así, al operar con los valores dados para las variables controladas por el marketing podemos comprobar la influencia sobre el resultado final del *CLV*. Esto nos permite, por ejemplo, identificar a los clientes más rentables o prever el *ROI* (Retorno sobre la inversión del inglés *Return Of Investment*) de los programas de fidelización, asignando los recursos del marketing de una forma eficiente y rentable para el negocio.

Su expresión aparece en la ecuación (4).

$$CLV = \sum_{t=0}^T \frac{(Mb_t - c_t)r_t}{(1 + i)^t} - AC \quad (4)$$

Los términos de esta expresión se definen a continuación:

Mb_t: Contribución al margen bruto del consumidor en el periodo *t*.

Para su cálculo resulta muy útil conocer la rentabilidad directa del producto adquirido por el cliente (*DPP* por sus siglas en inglés). El cálculo del *DPP* aparece en la Tabla 36:

| | |
|---|--|
| | Precio de venta |
| - | Coste de compra del producto. |
| = | Margen bruto |
| + | Condiciones especiales (rápel, comisiones, descuentos especiales, etc.) |
| = | Margen bruto ajustado |
| - | Costes directos del producto |
| = | Beneficio directo del producto |
| | Costes indirectos (gastos generales, costes de estructura, gastos financieros, amortizaciones, etc.) |
| = | Beneficio neto antes de impuestos |

Tabla 36. Cálculo del DPP

De este modo, si tenemos caracterizados a nuestros clientes, por ejemplo aquellos interesados por la dermoestética, de los que conocemos los productos que conforman su cartera base podremos calcular la rentabilidad de este perfil de clientes.

C_t : coste del servicio ofrecido al consumidor durante el periodo t .

Las empresas centradas en rentabilizar la relación con el cliente llevan a cabo estrategias de fidelización que repercuten en el *CLV* (por ejemplo tarjetas de fidelización). Además establecen contactos regulares con el cliente a través de los distintos canales (SMS, correo,...etc.). En definitiva, aquí se engloban los costes de publicidad y promoción.

i : tasa descuento.

La tasa de descuento es el tipo de interés utilizado para convertir una inversión futura en una cantidad actual. Está relacionada con los tipos de interés, con la inflación y con el periodo de evaluación considerado.

r_t : tasa de retención o probabilidad de que el cliente repita la compra.

Su valor se calcula analizando nuestra cartera activa de clientes entre dos periodos consecutivos. Por ejemplo si en un periodo dado disponemos en nuestra cartera de 100 clientes activos y en el periodo inmediatamente posterior pasan a 70 clientes, entonces la tasa de retención o probabilidad de que el cliente repita la compra es de 70/100, o del 70% si lo expresamos en porcentaje.

AC: coste de adquisición.

Se obtiene como el cociente entre el coste económico de la estrategia de captación y el número de clientes activados como consecuencia de la campaña desarrollada.

T: horizonte temporal de la predicción (número de meses, años...).

Es necesario que la tasa de descuento esté referida a esta unidad temporal (años, meses,...).

El *CLV* es muy útil para decidir si merece la pena abordar una campaña de fidelización o no en base a los parámetros vistos anteriormente. Por ejemplo, si pensamos abordar una campaña de 5 años con un coste anual inicial de 10.000, y 5.000 los restantes con los datos que aparecen en la Tabla 37, el *CLV* nos permite evaluar si merece o no la pena abordar dicha campaña de fidelización.

| | Sin campaña | Con campaña |
|--|----------------|----------------|
| Cientes activos en el año 1 | 1000 | 1000 |
| Valor de la venta por compra del cliente | 40 | 50 |
| Productos por cliente al año | 12 | 12 |
| Volumen de compra por cliente | 480 | 600 |
| Tasa de retención de clientes | 60% | 70% |
| % margen neto | 20% | 20% |

Tabla 37. Datos iniciales simulación CLV

Los resultados obtenidos con los datos de la Tabla 37 se recogen en la Tabla 38.

Comparativo de CLV con y sin campaña de Fidelización

| | Sin campaña | Con campaña |
|---|-------------|-------------|
| Año 1 | 96,00 € | 110,00 € |
| Año 2 | 149,83 € | 180,47 € |
| Año 3 | 180,02 € | 227,36 € |
| Año 4 | 196,94 € | 261,61 € |
| Año 5 | 206,44 € | 286,45 € |
| Con campaña (año 5) | | 286,45 € |
| Sin campaña (año 5) | | 206,44 € |
| Diferencia | | 80,02 € |
| Beneficios totales de la campaña | | 80.016,45 € |
| Mejora (%) del CLV aplicando una campaña | | 38,76% |

Tabla 38. Comparación del CLV con y sin campaña

Como puede observarse, en base a los datos iniciales, la simulación es favorable a realizar la campaña de marketing y fidelización puesto al llegar al año 5º, el CLV mejora los resultados un 38,76% realizando la campaña respecto de no hacerla. Por supuesto, si los resultados obtenidos hubiesen sido justo lo contrario, decidiríamos no realizar la campaña y así el propietario no incurriría en una inversión no rentable para la oficina de farmacia.

Centrándonos nuevamente en el concepto de fidelización, podemos definirlo como el conjunto de acciones encaminadas a conseguir relaciones estables y prolongadas con los clientes, de las que surgirán sentimientos positivos o algún tipo de compromiso de los clientes con la empresa (108).

El mayor valor de un plan de fidelización es conseguir crear un vínculo entre la empresa y sus clientes más rentables, así como conseguir retener a aquellos que, siendo solventes e interesantes para la empresa, muestran síntomas de propensión a la fuga.

Sin embargo, en algún sector más que otro, el mercado está saturado de planes y clubes que solo se basan en regalos y promociones. Un buen plan de fidelización debe basarse en una segmentación adecuada y en el uso de aplicaciones y soluciones de inteligencia

de clientes para ofrecer a cada uno de ellos los privilegios o ventajas acordes con su vinculación a la empresa. El concepto de segmentación lo trataremos en la siguiente sección y sucesivas de forma detallada.

Las actividades básicas que conforman un plan de fidelización son las siguientes (107):

1. Clasificar a los clientes: segmentación, definición de clientes prioritarios, seleccionar clientes de cartera.
2. Identificar grupos con los que trabajar: identificar los colectivos más interesantes, teniendo en cuenta la prioridad de los clientes y la tendencia o propensión al abandono y definir estrategias y acciones diferenciales para cada grupo según el atractivo del segmento en cuestión.
3. Identificar parámetros del plan: estrategia (mantenimiento o desarrollo de los clientes), objetivos a conseguir con el plan de fidelización, indicadores y metas.
4. Diseñar el plan de acción con iniciativas o acciones, los canales (más adecuados y rentables) así como el presupuesto destinado a cada acción.

En el caso de entorno de la oficina de farmacia, y según expertos del sector, la fidelidad del cliente depende, fundamentalmente, de dos factores (108):

- La satisfacción del cliente: un cliente/paciente que está satisfecho con el producto dispensado o con el servicio ofrecido, es un cliente fidelizado. Además, se debe tener en cuenta que un cliente fiel tiene un poder increíble de prescripción, por lo que se convierte en el “mejor comercial” de la oficina de farmacia.
- La competencia y el valor percibido de sus servicios: el cliente valorará los servicios de la oficina de farmacia con lo que se encuentra fidelizado, en comparación siempre con lo que la competencia le pueda ofrecer. Un cliente fidelizado mantiene la relación con nuestra farmacia sin evaluar cada vez que compra todas las posibles opciones que le ofrece la competencia. En el momento que la competencia consiga que el cliente fidelizado evalúe otras alternativas, su satisfacción más el coste de cambiar debe ser mayor que la valoración de la

competencia para mantenerlo fiel. La farmacia debe vigilar qué hace la competencia, para poder analizar sus puntos débiles y fuertes, para poder diferenciarse con su propia oferta.

Los beneficios para los clientes y las oficinas de farmacia de la fidelización son variados. A modo ilustrativo, encontramos algunos de ellos en la Tabla 39 (111).

| Paciente/Cliente | Farmacia |
|---|---|
| Ahorro de costes: cuando el paciente ha encontrado la farmacia que le aporta lo que necesita, ya no tiene que perder tiempo ni dinero en buscar otra farmacia | Los pacientes fieles generan más ingresos a la farmacia durante más tiempo |
| Disminución del riesgo: conoce perfectamente a su proveedor | Los costes de mantener a un paciente actual suele ser más bajos que la adquisición de otro nuevo |
| Mayor satisfacción de sus necesidades: los empleados de la farmacia le conocen y saben qué necesita | El paciente fidelizado tenderá a comprar el producto en nuestra farmacia y su lealtad aumentará a medida que gestionemos bien su fidelización |
| Reducción de errores en el uso y beneficio del producto, así como en la prestación del servicio: esto lo aporta el conocimiento mutuo | El paciente fidelizado compra más cosas cuando acude a nuestra farmacia que el ocasional |
| Consecución de un trato prioritario y personalizado | Nuestro paciente habitual está más abierto a comprar productos nuevos, que incluso hayamos desarrollado nosotros mismos, y ello facilita la venta cruzada de otros productos |
| Información, asesoramiento y colaboración por parte del farmacéutico | Un paciente satisfecho es la mejor publicidad para nuestro negocio. La comunicación boca-oreja es muy eficaz y ayuda a reducir los gastos que la farmacia tendría que destinar al marketing |
| | Al fidelizar a los pacientes generamos barreras protectoras y de entrada a la competencia (otras farmacias y/u otros establecimientos) |
| | Atender a nuestro paciente nos supone un ahorro en costes, porque al conocer mejor lo que quiere, cuesta menos atenderle bien |
| | Atender a un paciente fiel genera una mayor satisfacción y más rendimiento en nuestros empleados. Cuesta menos volcarse en los pacientes fieles que en los esporádicos, y ello repercute en la rentabilidad |
| | Los pacientes fieles asimilan mejor la subida de los precios porque perciben otros servicios |
| | Nuestros pacientes de siempre nos pueden aportar ideas para el desarrollo de nuevos productos y mejorar los servicios que les ofrecemos |

Tabla 39. Beneficios de la fidelización para clientes/pacientes y farmacias

En España, la fidelización en la farmacia ya ha dado sus primeros pasos; hoy es común el uso de tarjetas de fidelización en muchas de ellas. Cada vez más existe la conciencia que en la actualidad es necesario cubrir las expectativas de los clientes de forma efectiva. Existen un conjunto de factores relevantes (109) que debemos tener en cuenta antes de implantar planes y herramientas de fidelización:

- Evaluar nuestra clientela: averiguando los segmentos de clientes sobre los que es conveniente actuar y adoptar los métodos que más valor aporten a esos segmentos.
- Elección de los métodos de fidelización: estos estarán relacionados con la persona encargada de la toma de decisiones, habitualmente el farmacéutico titular, cuya mentalidad y forma de ver la profesión farmacéutica le llevará a decantarse por métodos más sanitarios o por métodos más comerciales.
- Los recursos económicos: ya que habrá que implantar aquellos métodos que representen una relación «valor para el cliente/coste para la farmacia» aceptable.
- El espacio: pues existen métodos cuya puesta en marcha se puede ver condicionada por los metros cuadrados con que cuente el local de la farmacia,
- El tiempo del que se dispone en la farmacia y el que requieren las diferentes acciones.
- El personal: tanto por su carga actual de trabajo como por su capacitación.

Respecto a la estructura de los planes de fidelización existen varias alternativas. Un ejemplo de ellas sería la siguiente (110):

- Estudiar el grado de satisfacción de la clientela actual.
- Analizar las preferencias de los clientes.
- Determinar el sistema de fidelización a emplear.
- Establecer mecanismos de recompensa.
- Informar a los clientes de la farmacia.

- Efectuar un control de resultados.

Por último, y gracias al estudio de *PM Tools Pharma* (112) en más de 400 farmacias, podemos conocer los cuatro frenos que declaran las oficinas de farmacia a la hora de abordar un plan de fidelización en las oficinas de farmacia:

- “No sé para qué me va a servir a mi fidelizar”.
- “No tengo tiempo de fidelizar”.
- “No tengo medios para fidelizar”.
- “Fidelizar es muy caro”.

Es evidente que el porcentaje de estos “frenos” difieren entre los farmacéuticos encuestados según sea su perfil (que vimos con anterioridad).

Con la situación expuesta en la primera parte de esta investigación acerca de la oficina de farmacia, es evidente que el mercado de los medicamentos no puede ser la única alternativa. La tendencia es, al igual que en Europa, elevar el porcentaje de parafarmacia, homeopatía, fitoterapia, servicios, etc. en los ingresos de la oficina de farmacia. Teniendo en cuenta la actual oferta en otros canales sólo cabe tener clientes informados, a los que hacerles llegar ofertas, descuentos, para que vuelvan. La pregunta no debería ser fidelizar o no, sino cuanto invierto en fidelizar a mis clientes.

Ya avanzamos al inicio de esta sección el estudio realizado por la oficina americana del consumidor que concluyó que cuesta cinco veces más conseguir un nuevo cliente que mantener un cliente actual; si a esto le sumamos el coste de recuperar un cliente que se ha marchado a otra oficina de farmacia, a un gran almacén, o a una web con descuentos, entonces la situación empeora notablemente. Como ya apuntamos en secciones anteriores, la farmacia no puede quedarse impasible en la situación actual de cambios y esta investigación le proporcionará elementos de reflexión y herramientas para su abordaje. Evidentemente, la fidelización es una de ellas.

Uno de los factores claves para el éxito de los planes de fidelización tiene que ver con el marketing focalizado, para lo cual, es fundamental contar con una adecuada segmentación. Tema principal de esta investigación y que abordamos a continuación.

1.3.3.2 Registro de pacientes y clientes

Antes de abordar el concepto de la segmentación, vamos a estudiar el registro de pacientes y clientes en las oficinas de farmacia puesto que es un requisito *sine qua non* para abordar la segmentación y por ende, cualquier plan o programa de fidelización.

Es preciso advertir que ninguna de las herramientas que hemos comentado con anterioridad tiene cabida en la oficina de farmacia si no tenemos registrada la información de nuestros clientes en bases de datos adecuadas. Habitualmente es el programa de gestión de clientes el responsable de configurar y mantener actualizadas las bases de datos, cuyo uso en las oficinas de farmacia está bastante consolidado.

Dichas bases de datos serán la fuente sobre las que utilizar herramientas que permitan conocer a nuestros clientes y aplicar análisis estadísticos para detectar segmentos de mercado y tratamientos diferenciados por tipologías de clientes. A partir de ellas se podrán planificar las estrategias y medir los resultados que las acciones estratégicas de la oficina de farmacia definan sobre sus clientes.

Según Manuel Ojeda (115) la evolución histórica del conocimiento de los usuarios en la oficina de farmacia ha pasado por varias fases:

- Empírica.
- Manual.
 - Comercial: registro de créditos pendientes. (Clientes-Provisional).
 - Profesional: libro recetario y de estupefacientes (Pacientes - Permanente).
 - Atención farmacéutica: se registran para los pacientes sus patologías y tratamientos. (Pacientes-Permanente).
- Informatizada.
 - Comercial: registro de créditos pendientes a través de los programas de gestión (Clientes-Permanentes).
 - Profesional: cumplimentación del libro recetario a través de los programas de gestión (Pacientes-Permanente).

- Atención Farmacéutica: a través de los programas de atención farmacéutica del Bot Plus (Pacientes-Permanente).
- Receta electrónica: a través de los programas de gestión. No registra pacientes, tratamientos y patologías. Solo informa del tratamiento de los pacientes. (Pacientes-Permanente).

El objetivo del registro de los datos sobre los clientes de la oficina de farmacia es disponer de datos estructurados y normalizados sobre todo aquello que permita darle el mejor servicio en base a la actuación profesional del farmacéutico. Manuel Ojeda (115) apunta en este sentido, los siguientes grupos de información a recoger sobre los clientes:

- Datos del cliente.
- Beneficiarios.
- Hábitos.
- Servicios.
- Salud.

En el Anexo 6.1 puede verse el cuestionario para la recogida de información sobre estos cinco bloques y que ha sido implementado informáticamente por el autor de esta investigación en la oficina de farmacia de Manuel Ojeda situada en Sevilla.

Una buena base de datos que permita realizar un buen análisis sobre los clientes de una oficina de farmacia debe contener, al menos, la siguiente información:

- Historial de compras y tipo de productos adquiridos.
- Información descriptiva sobre quien es el cliente, datos demográficos y un historial de los posibles incidentes que han existido en la relación cliente/paciente.
- Análisis por cliente de los productos que, en función de su perfil, debe consumir pero que no compra en nuestra farmacia.

La base de datos se convertirá en el componente primordial del plan de fidelización, proporcionando la información sobre datos para calcular el *CLV* y sobre como incrementar el número de clientes, el valor que cada uno de ellos aporta y la estrategia de comunicación adecuada para cada segmento de clientes.

1.3.3.3 Segmentación de clientes

Desde luego, y los estudios aquí mostrados así lo resaltan, si una característica ha destacado de la farmacia desde sus inicios ha sido el trato personalizado prestado por el farmacéutico. Hecho mostrado por las encuestas y refrendado por la tendencia de los pacientes a acudir a su oficina de farmacia de “referencia” de forma recurrente. No obstante, y sobre todo en el mercado del autocuidado de la salud y de los servicios farmacéuticos complementarios que se están comenzando a definir ahora, el número de productos (cada vez más sofisticados) y las necesidades particulares de los consumidores (que se cuentan por miles) arrojan un número de combinaciones posibles que hasta el farmacéutico con mayor capacidad de memoria sólo guardaría la información de un pequeño grupo de clientes. Además, aunque fuese capaz de tener en su memoria toda la información, no podría mantener una relación personalizada con todos y cada uno de los clientes ni tendría recursos suficientes para ello. Sin embargo, hoy nos encontramos frente a un nuevo paradigma del marketing (114) por el que la principal fuente de ventajas competitivas sostenibles para la empresa está basada en el conocimiento en profundidad de los deseos y demandas de cada uno de los consumidores.

En este sentido, la segmentación puede representar una solución de compromiso a este dilema. Por tanto, la segmentación de mercados como concepto viene a significar, esencialmente, la identificación de consumidores como miembros de grupos que tienen en común hábitos de compra y uso de productos y servicios similares, y que responden de manera parecida a las acciones del marketing.

Las empresas españolas hace ya tiempo que saben que no basta con aumentar las ventas o lograr nuevos clientes. En tiempos en los que internet y las redes sociales permiten a

los ciudadanos tener mucha información acerca de las distintas ofertas que ofrecen las compañías, se hace imprescindible conocer al consumidor, ofrecerle lo mejor para él y hacer que se sienta satisfecho e incluso adelantarse a sus necesidades y deseos (112). La segmentación de clientes puede ser de gran utilidad para ofrecer a cada cliente justamente lo que necesita.

Edwar H. Chamberlin (118) y Joan Robinson (119), estudiaron la noción de competencia monopólica y la teoría de competencia imperfecta, nociones ambas basadas en la idea de mercados heterogéneos. Partiendo de esta idea, Wendell R. Smith introdujo en 1956 la noción de segmentación de mercados por primera vez en la literatura especializada en marketing (115), un concepto que habría de convertirse en una idea fundamental tanto en la teoría como en la práctica del marketing. Fue Smith quien acuñó la primera definición de segmentación en la página seis de su artículo: “dada la heterogeneidad de las demandas de los consumidores, la segmentación del mercado implica la definición de un número no determinado de submercados homogéneos que responden a preferencias distintas, producto de la demanda de los consumidores hacia una satisfacción más precisa de sus deseos” (116).

La existencia de distintos segmentos de consumo, tanto en mercados de consumidores finales como en mercados industriales, ha dado pie dentro de las empresas a la concepción de la estrategia de posicionamiento mediante la diferenciación. Es decir, toda empresa, tomando en consideración los recursos financieros, técnicos, físicos y humanos con los que cuenta, debe decidir a qué segmentos de mercado debe acudir para ofrecer sus productos y servicios, evaluando el tamaño y capacidad de compra de los mismos (rentabilidad), y la posibilidad de que su oferta sea percibida como distinta por estos mismos consumidores (posicionamiento) gracias a la utilización de un *marketing mix* diferente. Por tanto, las empresas segmentan sus mercados para capturar con mayor eficacia las demandas y deseos de grupos de consumidores potenciales, así como para conservar a todos aquellos consumidores que se han convertido ya en clientes.

La segmentación de mercados puede definirse como el proceso de dividir el mercado en grupos uniformes más pequeños, que tengan características y necesidades semejantes. Este hecho deriva del reconocimiento de que el mercado está conformado por subgrupos llamados segmentos. Dichos segmentos son grupos homogéneos de consumidores que responden de la misma forma a un conjunto de estímulos y variables. Por ello, su búsqueda es un objetivo básico de las estrategias de marketing. Está búsqueda dentro de los clientes de una oficina de farmacia es uno de los objetivos principales de esta investigación y por ello lo trataremos de forma adecuada.

Las principales razones por las cuales la segmentación es necesaria (120) se muestran a continuación:

- Identifica los clientes más y menos rentables.
- Permite focalizar el marketing en los clientes, como manera de obtener mayor eficiencia en las campañas.
- Evita los mercados que no son rentables para la empresa.
- Construye una relación de lealtad con el cliente por medio de ofertar productos y servicios que estos quieren.
- Permite mejorar el servicio al cliente.
- Permite adelantarse a la competencia en productos y servicios específicos.
- Permite usar los recursos en forma eficiente.
- Permite mejorar la oferta en función de las necesidades del cliente.
- Permite incrementar las ganancias por medio de la disminución de costos y el aumento en precio que se le pueden dar a algunos productos o servicios.

Según Valderrey (117) y Ferrel y Hartline (124) los requisitos para conseguir una buena segmentación son los siguientes:

- Homogeneidad en el segmento.

Los miembros deben ser parecidos entre sí respecto de las características que los agrupan como segmento.

- Heterogeneidad entre segmentos.

Los segmentos deben ser diferentes entre sí respecto de las características que han conformado los segmentos.

- Estabilidad entre segmentos.
- Los segmentos deben ser identificables y medibles.

Las características de los miembros del segmento deben ser fácilmente identificables. Esto permite a la empresa medir las características de identificación, incluido el tamaño del segmento y el poder de compra

- Los segmentos deben ser accesibles y manejables.

El segmento debe ser accesible en términos de comunicación (publicidad, correo, teléfono, etc.) y distribución (canales, comercializadores, establecimientos minoristas, etc.).

- Los segmentos deben ser lo suficientemente grandes como para ser rentables.

El segmento debe ser lo suficientemente grande y rentable a efecto de hacerlo valioso para la empresa. El potencial de utilidades debe ser mayor que los costos implicados en la creación de un programa de marketing específicamente para el segmento.

- Los segmentos deben ser viables y sostenibles.

Los segmentos deben cumplir con el criterio básico de intercambio; incluido estar listos, dispuestos y poder realizar negocios con la empresa. También deben ser sostenibles en el tiempo para permitir a la empresa desarrollar eficazmente una estrategia de marketing a efecto de atender sus necesidades.

- Los segmentos deben ser sensibles.

Los segmentos deben responder a los esfuerzos de marketing de la empresa. Incluidos cambios en el programa de marketing tras el paso del tiempo. También deben responder de forma diferente a los otros segmentos.

En la Tabla 40 mencionamos los criterios, bases o variables para segmentar el mercado de consumo y ejemplos de dichas variables (124).

| Categoría | Variables | Ejemplos |
|---------------------------|-------------------------|--|
| Segmentación por conducta | Beneficios de producto | Calidad, valor, gusto, mejoramiento de la imagen, belleza, deportividad, velocidad, emoción, entretenimiento, nutrición, comodidad |
| | Uso del producto | Usuarios fuertes, medianos y ligeros; no usuarios; anteriores usuarios; usuarios de primera vez |
| | Ocasiones o situaciones | Emergencias, celebraciones, cumpleaños, aniversarios, bodas, nacimientos, funerales, graduaciones |
| | Sensibilidad al precio | Sensible al precio, consciente del valor, consciente del estatus (no sensible al precio) |
| Segmentación demográfica | Edad | Recién nacidos, 0-5; 6-12; 13-17; 18-25; 26-34; 35-49; 50-64; más de 65 |
| | Sexo | Hombre, mujer |
| | Ingresos | Menos de 15 000€; 15 000€-30 000€; 30 000€-50 000€; 50 000€-75 000€; 75 000€-100 000€; más de 100 000€ |
| | Ocupación | Obreros, empleados administrativos, técnicos, profesionales, gerentes, peones, jubilados, amas de casa, desempleados, etc. |
| | Educación | Sin estudios, Estudios primarios o equivalentes, Enseñanza general secundaria, 1er ciclo, Enseñanza Profesional de 2º grado, 2º ciclo, Enseñanza general secundaria, 2º ciclo, Enseñanzas profesionales superiores, Estudios universitarios o equivalentes |
| | Ciclo de vida familiar | Soltero, casado sin hijos, casado con hijos pequeños, casado con hijos adolescentes, casado con hijos grandes, divorciado, viudo, separado |
| | Generación | Generación Y, generación X, baby boomers, tercera edad |
| | Grupo étnico | Caucásico, afroestadounidense, latino, asiático |
| | Religión | Protestante, católico, musulmán, hindú, etc. |
| | Nacionalidad | Estadounidense, español, japonés, australiano, coreano, etc. |
| | Clase social | Alta, media, baja, clase trabajadora, nivel de pobreza |
| Segmentación psicográfica | Personalidad | Extrovertido, tímido, compulsivo, individualista, materialista, de mente cívica, ansioso, controlado, aventurero, etc. |
| | Estilo de vida | Entusiasta de los deportes al aire libre, mente deportiva, muy casero, adicto a la televisión, centrado en la familia, adicto al trabajo, etc. |
| | Motivaciones | Seguridad, estatus, relajamiento, comodidad, etc. |
| Segmentación geográfica | Regional | Noreste, sureste, medio oeste, Nueva Inglaterra, sur de Francia, Sudáfrica |
| | Tamaño de ciudad o país | Menos de 50 000; 50 000-100 000; 100 000-250 000; 250 000-500 000; 500 000-1 000 000; más de 1 000 000 |
| | Densidad de población | Urbana, suburbana, rural |

Tabla 40. Variables de segmentación de mercados utilizadas en los mercados de consumo

Según Luis Godás (123), la segmentación del mercado debe llevarse a la práctica de tal forma que permita desempeñar una estrategia comercial diferenciada que conduzca a la satisfacción de forma efectiva de las demandas de los consumidores. En este sentido, la utilidad de la segmentación de mercado pasa por conocer las características, deseos y necesidades de los consumidores (clientes/pacientes), así como sus actitudes y preferencias, con el objeto de poder diseñar estrategias comerciales adecuadas para atenderlos de forma rentable para ambas partes.

Una vez que una empresa ha decidido utilizar la segmentación de mercados y ha terminado de segmentar un mercado, debe evaluar cada segmento para determinar su atractivo y si ofrece oportunidades que concuerden con sus capacidades y recursos. Disponer de un segmento bien identificado y con las propiedades y características comentadas en esta sección, no significa que la empresa deba seguirlo. Aun así, un segmento atractivo puede abandonarse por varias razones: la falta de recursos, la no alineación con la misión de la empresa, una competencia feroz, un cambio inminente en la tecnología o las preocupaciones éticas o legales sobre identificar como meta algún segmento en particular. Pudiendo tener esta última razón un cierto impacto en el mercado de las oficinas de farmacia en el que nos focalizamos.

Previamente a realizar el diseño y la aplicación de una estrategia comercial concreta es necesario determinar qué prioridades tiene la empresa (la oficina de farmacia en nuestro caso) y en qué mercados/segmentos quiere estar presente. Una vez realizado este análisis, la empresa tiene a su disposición diferentes tipos de estrategias básicas, así como instrumentos comerciales con que llevarlas a cabo.

En la Tabla 41 aparecen diversas estrategias, según Kotler (121), que puede seguir una empresa a la hora de seleccionar un mercado objetivo o meta y poder así atender sus necesidades.

| Enfoque de un solo segmento | | | | | |
|-----------------------------|----|----|----|----|----|
| | M1 | M2 | M3 | M4 | M5 |
| P1 | | | ● | | |
| P2 | | | | | |
| P3 | | | | | |
| P4 | | | | | |
| P5 | | | | | |

| Enfoque selectivo | | | | | |
|-------------------|----|----|----|----|----|
| | M1 | M2 | M3 | M4 | M5 |
| P1 | | ● | | | |
| P2 | | | ● | | |
| P3 | ● | | | | |
| P4 | | | | | ● |
| P5 | | | | ● | |

| Especialización del mercado | | | | | |
|-----------------------------|----|----|----|----|----|
| | M1 | M2 | M3 | M4 | M5 |
| P1 | | | ● | | |
| P2 | | | ● | | |
| P3 | | | ● | | |
| P4 | | | ● | | |
| P5 | | | ● | | |

| Enfoque de mercado masivo | | | | | |
|---------------------------|----|----|----|----|----|
| | M1 | M2 | M3 | M4 | M5 |
| P1 | ● | ● | ● | ● | ● |
| P2 | ● | ● | ● | ● | ● |
| P3 | ● | ● | ● | ● | ● |
| P4 | ● | ● | ● | ● | ● |
| P5 | ● | ● | ● | ● | ● |

| Especialización del producto | | | | | |
|------------------------------|----|----|----|----|----|
| | M1 | M2 | M3 | M4 | M5 |
| P1 | ● | ● | ● | ● | ● |
| P2 | | | | | |
| P3 | | | | | |
| P4 | | | | | |
| P5 | | | | | |

Tabla 41. Estrategias básicas de una empresa para seleccionar el mercado meta. P=producto/servicio, M=Mercado

El propio Luis Godás (123), propone un conjunto de recomendaciones prácticas a la hora de conjugar las herramientas y técnicas que tienen como base el uso de la segmentación:

- Formar grupos de clientes que presenten un comportamiento de compra similar o un perfil homogéneo.
- Determinar el grado hasta el que cada segmento puede ser atractivo para la compañía.
- Valorar la rentabilidad de los segmentos considerados como atractivos.
- Establecer de qué forma se puede incidir sobre cada segmento y poner en marcha la estrategia adecuada.
- Combinar de forma acertada los instrumentos comerciales disponibles: producto, precio, distribución y promoción.

A estas recomendaciones habría que añadir una actividad de seguimiento y control de la marcha del plan y de la vigencia de los segmentos identificados.

En el siguiente apartado, abordamos la segmentación desde el punto de vista estadístico, analizando con detalle su conceptualización y las técnicas disponibles para su realización.

1.4 El problema estadístico de la segmentación

1.4.1 Introducción

El problema del análisis de la información y las bases de datos ha generado históricamente un gran interés. Pero cierto, es ahora, en los últimos años, con la aparición de las redes sociales y otras tecnologías, cuando el tema está concentrando la atención de grandes empresas, tecnólogos, matemáticos, ingenieros, etc. El sector farmacéutico no es ajeno a ello.

Actualmente, se disponen de multitud de herramientas para conseguir y almacenar miles de datos relacionados con nuestra empresa y los clientes. Existen muchos canales de comunicación por los cuales se pueden obtener información de utilidad. Generalmente, se han obtenido a través de encuestas, presenciales o no, a los clientes. Pero hoy en día, con la irrupción de las nuevas tecnologías, cada vez que nos conectamos a Internet buscando información, cada vez que accedemos a una aplicación en el móvil, que pagamos en una tienda y acumulamos puntos con la tarjeta de fidelización, cada vez que subimos al metro, etc., generamos información que bien analizada puede servir para conocer los hábitos y tendencias de consumo, el coche que me gusta, la playa que me gustaría visitar, el partido al que podría votar, la enfermedad que podría padecer, mis hábitos de desplazamiento y mucho más.

Y la cuestión principal es ¿qué podemos hacer con toda esta información? ¿Es posible analizarla? ¿Qué información podemos extraer? ¿Podemos aprender algo del histórico para conocer mejor a nuestra empresa y nuestros clientes? ¿Existen patrones de comportamiento “ocultos” en nuestros datos? Y es más, ¿podemos usar toda esta información para predecir el futuro? ¿Podemos definir nuevos productos y servicios adaptados a las necesidades de los clientes?

Ciertamente, la respuesta es positiva para todas ellas. Analizando las bases de datos apropiadamente, filtrando y eliminando la información irrelevante, seleccionando las variables apropiadas y, sobre todo, aplicando las técnicas estadísticas adecuadas es posible obtener resultados antes inimaginables.

Existen multitud de casos de éxitos donde los modelos econométricos, la inferencia bayesiana y la minería de datos están siendo usados para explicar y entender la dinámica de las empresas. Véase el caso de Coca-Cola (127), las múltiples aplicaciones de Nate Silver para predecir las elecciones presidenciales en EE.UU. o los resultados de la NBA o de *baseball* (120), el uso de segmentación para predecir una enfermedad (121), la clasificación del riesgo crediticio o la lucha contra el fraude (122).



Figura 53. Big Data

Para dar respuesta a estas y otras cuestiones surgen nuevos campos de investigación como *Big Data*, *Cloud* o *Cloud Computing*, que conjugan una de las ramas clásicas de la Estadística y la Investigación Operativa como es la Minería de Datos (en inglés, *Data Mining*) con la irrupción de la Tecnología de la Información, Internet, la redes sociales y las supercomputadoras.

Son muchas las aplicaciones reales del *Big Data y Cloud Computing* y se estima que aquellas titulaciones universitarias que se especialicen en ellas alcanzarán unas tasas de paro cercanas al 0%. En los últimos años han surgido estudios universitarios, y sobre todo másteres, en estas disciplinas enfocados para matemáticos, ingenieros e informáticos que deseen especializarse en este campo (123).

Estas nuevas líneas de investigación se ramifican a su vez en otras según la especialización del tema y los objetivos concretos que se pretenden lograr. Así, en función de la estructura de los datos se puede, por ejemplo, intentar predecir si un paciente padece una enfermedad o no a partir de los síntomas de otros pacientes con y sin dicha enfermedad; si un cliente o un perfil de cliente comprará o no en tu farmacia cuando se le informe sobre una promoción; si existen más o menos perfiles de clientes en tu negocio a los que diseñarles diferentes campañas publicitarias; si caerá o no la demanda de un producto al variar su precio; si ganará o no las elecciones un candidato a partir del tráfico de mensajes en las redes sociales que lo nombran, y muchas otras situaciones más.

Analizando con detenimiento estos y otros ejemplos, son muchas las clasificaciones y ramificaciones que se pueden proponer para los problemas de segmentación:

- Según se desee optimizar un objetivo o varios simultáneamente distinguiremos entre problemas mono-objetivos y multi-objetivos.
- Según se desee predecir o no.
- Según se disponga de una base de datos grande o no.
- Según se puedan aplicar técnicas de optimización exacta o, por la complejidad del problema, se requieren de algoritmo aproximados.
- Según los datos estén “etiquetados” o no hablaremos de problemas de clasificación o de búsqueda de conglomerados (segmentación).

Justamente es este último caso el que nos ocupará en este capítulo. Como ya se ha expuesto anteriormente, uno de los objetivos principales de la Tesis es proporcionar a la oficina de farmacia una herramienta eficaz con la que conocer los distintos perfiles de sus pacientes y/o clientes y realizar una correcta toma de decisiones estratégicas para ésta. No obstante, otros trabajos futuros para seguir profundizando y que seguramente podrían ser objeto de otras tesis doctorales serían: la predicción de ventas en la farmacia, la optimización del stock de una farmacia, el análisis dinámico de ambos problemas, etc.

1.4.2 Segmentación frente a clasificación

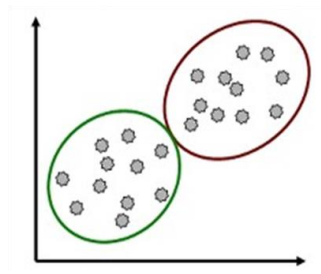
En esta sección, se explicará la diferencia entre segmentación y clasificación con idea de aclarar al lector que la aplicación en la que nos centramos en esta disertación se engloba dentro de lo que conocemos como problema de segmentación o conglomerados.

- En un problema de clasificación o de aprendizaje supervisado se parte de un conjunto de elementos y una serie de grupos ya predefinidos o existentes a los que pertenecen dichos elementos. A partir de esta información se pretenden inferir reglas, patrones o ecuaciones que clasifiquen exitosamente a futuros elementos. Generalmente, el conjunto de elementos se divide en dos: un conjunto de entrenamiento o aprendizaje y un conjunto de prueba o verificación. A partir del primer conjunto se obtienen los patrones y con el segundo conjunto se evalúa la robustez de éstos.
- En un problema de segmentación o conglomerados (o de aprendizaje no supervisado) se agrupan los elementos por similitudes (generalmente distancias). El número de grupos es generalmente desconocido, no se requiere de un conjunto de entrenamiento y no se puede comprobar su fiabilidad con otro conjunto.

En la Figura 54, el color rojo y verde de los puntos de la figura de la derecha corresponde a la clásica situación de un problema de clasificación. Tras aplicar una de las muchas técnicas de clasificación (por ejemplo, análisis discriminante) se obtiene una recta que los divide. Por el contrario, en la figura de la izquierda todos los puntos tienen el mismo color y se agrupan en varios grupos según las distancias entre ellos o algún otro criterio. El número de grupos no suele estar prefijado y supone una complejidad extra al problema. Cada grupo suele representarse por su centroide y futuros elementos serán asignados al grupo más cercano, esto es, al grupo con centroide más cercano.

SEGMENTACIÓN

- Datos no etiquetados
- Los puntos en cada grupo están cerca unos de otros
- Se identifican estructuras o patrones en los datos
- Aprendizaje no supervisado

**CLASIFICACIÓN**

- Datos etiquetados
- Se busca una regla que asigne las etiquetas a nuevos puntos
- Aprendizaje supervisado

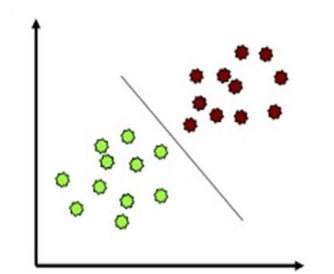


Figura 54. Diferencias entre segmentación y clasificación

Un ejemplo clásico de clasificación es la medición del riesgo en banca. Se dispone de una base de datos de los clientes con información relevante que puede explicar su disposición a pagar o no un crédito (importe del crédito, ingresos, edad, otros créditos...) y una variable binaria que los clasifica en dos grupos: los que han pagado créditos anteriores y los que no. Claramente, es de un gran interés para la empresa conocer las “reglas” que explican un impago para aplicárselas a futuros clientes y predecir si pagará o no.

En el caso de la segmentación, las aplicaciones más comunes se dan en el sector del marketing. En estos casos, lo habitual es dividir a los clientes en grupos con similares características relevantes para las ventas de la empresa que ya vimos en la Tabla 40 (edad, frecuencia de visita, gasto mensual, hábitos de consumo...). El número de perfiles o grupos de clientes es desconocido y no tiene por qué ser común entre las empresas.

En esta investigación, utilizaremos indistintamente los términos “segmento”, “clúster”, “grupo” o “conglomerado” para referirnos al conjunto de elementos que poseen características similares. El uso de uno u otro término dependerá de la sección en que se trate, por aquello de aplicar el término más utilizado en la literatura existente.

1.4.3 Segmentación de bases de datos. Pros y contras

Como hemos comentado anteriormente, las técnicas de segmentación tratan de organizar o agrupar a un conjunto de objetos en grupos, generalmente disjuntos, con ciertas similitudes. Así pues, simultáneamente cada grupo consistiría de elementos similares entre sí y los más diferente posible a los elementos de los otros grupos. Recomendamos dos *surveys* de gran calidad a todo aquel que desee profundizar en este problema y sus variantes. Son los trabajos de Xu y Wunsch (124) y Berkhin (125).

Este proceso es muy natural e intuitivo para el ser humano y resulta de gran interés para categorizar y clasificar la información en grupos con elementos o características comunes. En la Figura 55 se expone un ejemplo en 2 dimensiones. A simple vista se aprecian 3 grupos y únicamente dudamos con los puntos “frontera”, esto es, aquellos que están aproximadamente a la misma distancia de un par de grupos.

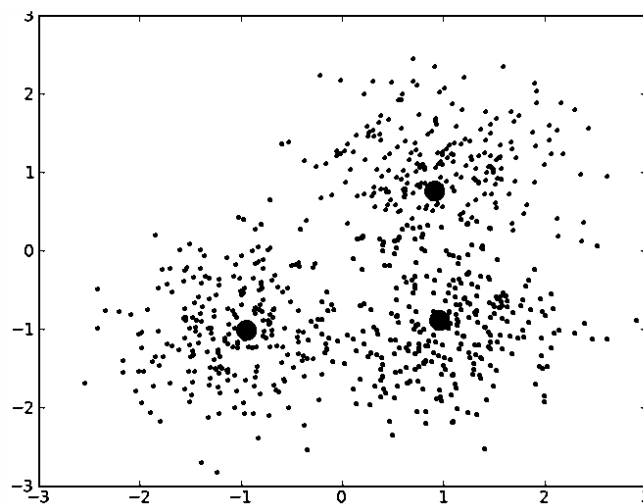


Figura 55. Nube de puntos en dos dimensiones

Una vez definido el criterio de similitud (por ejemplo, la distancia euclídea) y aplicada algunas de las técnicas de segmentación, se obtendría una solución como la siguiente.

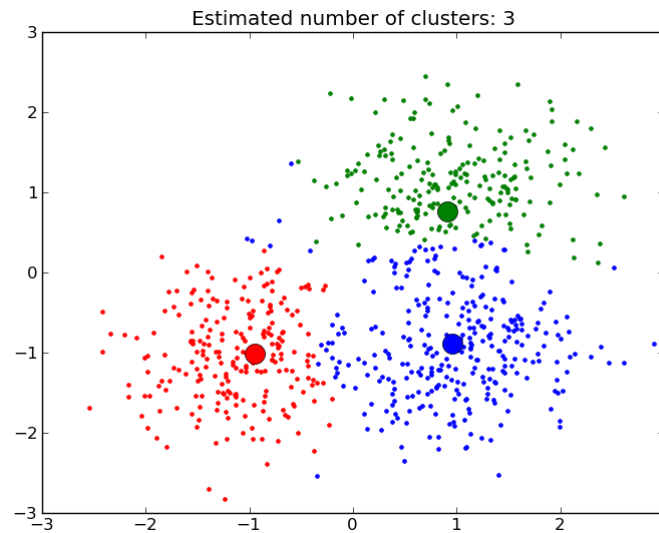


Figura 56. Ejemplo de tres clúster

No obstante, así como esta tarea puede ser sencilla para un ser humano teniendo en cuenta la complejidad, capacidad y velocidad de procesamiento del cerebro, la complejidad para resolver este tipo de problemas para una máquina es mucho mayor. Por este motivo, estos problemas han centrado la atención de científicos y empresarios en los últimos años.

Pero a pesar de que la mejor máquina existente es el cerebro, el problema de la abundancia de datos de los que se dispone actualmente hace imposible el poder manejarla correctamente. El problema es aún mayor si se tiene en cuenta que cada objeto o elemento puede estar caracterizado por un gran número de atributos o variables que, a su vez, pueden ser numéricas o no. Así como la representación gráfica de la información para 2 ó 3 variables es “comprensible” y visualmente analizable, esta tarea se hace imposible para más de 3 atributos. Por tanto, es necesario generalizar este concepto, intuitivo para 2 ó 3 dimensiones, a cualquier dimensión.

Por todo lo anterior, la segmentación de la información es un problema tratado históricamente desde muchos enfoques: teoría de grafos, estadística, inteligencia artificial, redes neuronales, conjuntos difusos... En las próximas secciones revisaremos y clasificaremos estas técnicas para centrarnos en la usada para nuestro problema: optimización matemática.

Como acabamos de comentar, formularemos nuestro problema desde el punto de vista de la optimización. Esto es, a partir de ciertas medidas de similitud y otras medidas estadísticas, se buscará la partición de la información que optimice (minimizando o maximizando) la optimización mono-objetivo o la optimización multi-objetivo. Las funciones objetivo son generalmente definidas como medidas estadísticas (como por ejemplo, la varianza) que reflejan el grado de similitud entre los elementos de un mismo grupo y/o el grado de diferencia entre los que no son del mismo grupo. En siguientes subsecciones introduciremos alguna de las funciones más utilizadas.

Los pros y los contras de este tipo de planteamiento son los siguientes. Como ventaja principal está la simplicidad en el planteamiento del problema. Al formular el problema de segmentación como un problema de optimización mono o multi-objetivo, es posible usar multitud de modernas técnicas de optimización ya consolidadas. Existen en la actualidad técnicas de optimización exactas que garantizan encontrar la solución al problema (aunque requieren de mucho esfuerzo computacional y solo son válidas para problemas pequeños) o heurísticas y metaheurísticas que en tiempos muy inferiores obtienen soluciones muy próximas a la óptima (generalmente para problemas de tamaño medio-grande). En esta tesis, debido al tamaño del problema, nos centraremos en el uso de las técnicas metaheurísticas específicamente diseñadas para problemas de segmentación.

Además, dependiendo de la función objetivo usada, la o las soluciones están claramente determinadas y definidas por éstas.

Entre las principales desventajas destacamos la facilidad de muchos de los algoritmos para quedar atrapados en óptimos locales, la enorme dependencia de la solución de la medida de similitud utilizada (por lo que pueden existir muchas soluciones a un mismo problema) o el no conocimiento del número de clústeres con antelación.

En la Figura 57 mostramos distintas geometrías de conglomerados. En función de la medida de distancia o similitud utilizada o el algoritmo escogido se pueden obtener soluciones con 2, 3 y hasta 4 grupos, que detectan o no las distintas estructuras de los

datos. Insistimos en que todas ellas son soluciones óptimas del problema según la función de similitud usada o el número de grupos deseados.

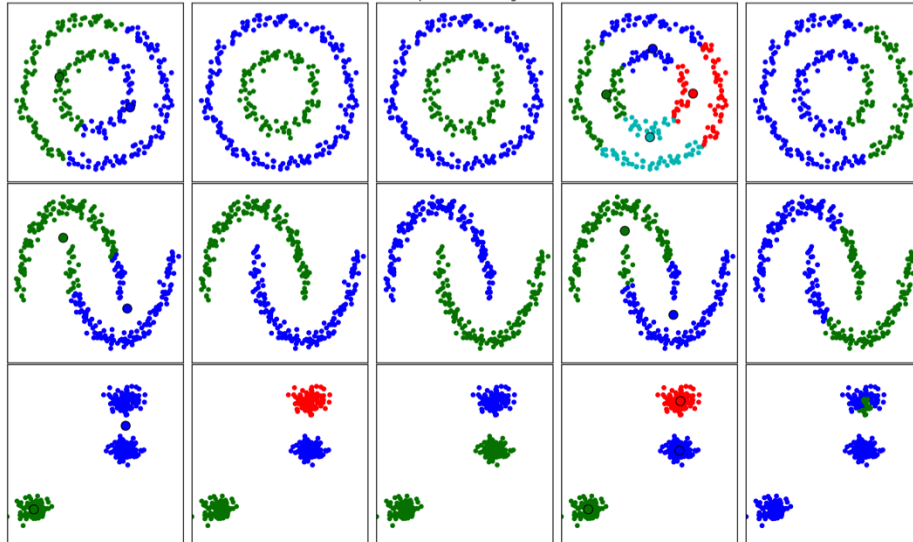


Figura 57. Diferentes geometrías de conglomerados

En las próximas secciones ahondaremos en estos puntos.

1.4.4 Determinación del número de clústeres

Estimar el número de clústeres en una base de datos es una tarea muy compleja dentro de los problemas de segmentación. Como ya hemos comentado, la segmentación se clasifica dentro del *aprendizaje no supervisado*, y se asume que los algoritmos deben tener la capacidad de determinar el número de grupos existentes, en la práctica la mayoría de los algoritmos consideran este número como un *input* (este es el caso, por ejemplo, del algoritmo *K-means* y sus variantes) ayudando a converger al algoritmo al óptimo para el valor dado.

Es más, determinar el número de clústeres óptimo es imposible sin cierta información previa. Como se apunta en Das, Abraham y Konar (134), existen multitud de situaciones reales donde es imposible estimar este valor sin haberlo analizado previamente. Este podría ser el caso de la segmentación de textos de un idioma desconocido, o de documentos tras una consulta a un buscador, etc.

Hasta la fecha no existe un método óptimo para encontrar el número grupos presentes en los datos. En esta sección, revisaremos los principales métodos propuestos para determinar este número. Como se verá, los más populares se basan en el uso de indicadores de validez de la segmentación (*Clustering Validity Indices, CVI*) que asignan un valor a cada solución del problema obtenida en función de distintos criterios (número de clústeres, compacidad, distancia entre clúster, etc.). Para ellos se usan multitud de funciones o distancias entre elementos y medidas estadísticas derivadas de éstas (como la varianza por ejemplo).

El problema es que los CVI suelen estar definidos con un fin específico: compacidad, geometría, densidad, etc., de los clústeres. Por tanto, son robustos en su criterio pero ignoran otras de las propiedades. En definitiva, no existe una única medida o indicador capaz de tener en cuenta todas las características requeridas y suele ser la combinación o uso de varias de ellas simultáneamente una alternativa factible al problema.

Dos revisiones de la literatura muy recientes sobre el tema son los trabajos de Dimitriadou y col. (127) para datos binarios, y de Vendramin y col. (128) que revisan los principales medidas basadas en indicadores.

1.4.4.1 Regla general

Una de las primeras reglas para estimar el número adecuado de grupos es la siguiente:

$$k = \sqrt{\frac{n}{2}}$$

donde n el tamaño de la base de datos y k es el número de clústeres buscados. En la Tabla 42 mostramos el valor de k obtenido por este método para distintos tamaños de datos.

| n | k (sin redondear) |
|------|----------------------|
| 50 | 5,00 |
| 100 | 7,07 |
| 500 | 15,81 |
| 1000 | 22,36 |
| 2000 | 31,62 |
| 5000 | 50,00 |

Tabla 42. Valores de K para diferentes tamaños de volumen de datos

Como se ha comentado anteriormente, no existen estudios que determinen una regla general para determinar el número de clústeres. Es más, no existe una medida o función capaz de reflejar las distintas geometrías y características de todas las bases de datos. Por tanto, la única ventaja de esta regla es su simplicidad mientras que son muchas sus desventajas: la no dependencia de la dimensionalidad de la base de datos (número de atributos) o la independencia de la fórmula respecto a la estructura geométrica de los datos.

1.4.4.2 El método del codo (*Elbow method*)

El método del codo mide el porcentaje de la varianza explicada por los grupos en función del número de grupos. Teniendo en cuenta que la pendiente de esta función no decreciente es elevada en los primeros valores del número de clústeres y a medida que aumenta el número éste crece lentamente, el valor donde se produce este cambio de tendencia y la ganancia marginal podría corresponder al óptimo número de grupos.

A continuación, mostramos un ejemplo de la literatura donde el punto de inflexión en la curva corresponde al valor óptimo para el número de clústeres.

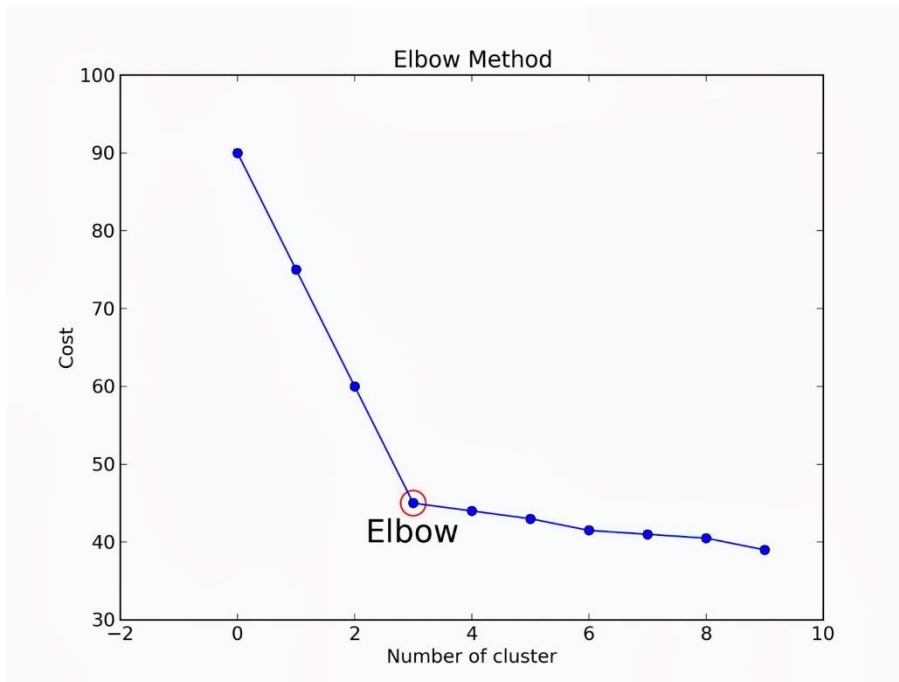


Figura 58. Método de Elbow

En la Figura 58 se aprecia el cambio de tendencia a simple vista pero esto no es necesariamente así, lo que le infiere un carácter subjetivo al método. En Tibshirani y col. (129) solventan este problema mediante el uso de herramientas estadísticas que determinen unívocamente dicho valor.

1.4.4.3 Indicadores

Otra de las técnicas más habituales se basa en la optimización de una función o indicador de coste, aptitud o validación. Dicho indicador suele considerar el número de clústeres y diversas medidas de error o similitud para medir la validez de una solución dada. El óptimo (máximo o mínimo) de dicha función se alcanza con el número óptimo de clúster. En Vendramin y col. (128) se revisan los principales indicadores de validez para problemas de segmentación. Otra revisión de éstos puede verse en Desgraupes (138). Veamos uno de los más conocidos a modo de ejemplo y dejemos para la siguiente sección una revisión de los más utilizados.

Como ejemplo puede citarse el de *Calinski-Harabasz* (130) cuyo índice se define como:

$$CH = \frac{\text{traza}(S_B) N_p - 1}{\text{traza}(S_W) N_p - k}$$

donde S_B y S_W son las matrices entre e intra-cluster, N_p el número de puntos y k el número de clústeres. Recordemos que la traza de una matriz se define como la suma de los elementos de su diagonal principal. El valor máximo del indicador de Calinsky-Harabasz determina la solución óptima.

En la Figura 59 se muestra un ejemplo de la evolución del indicador de Calinsky-Harabasz. El máximo del indicador en este caso se alcanza con 5 clústeres.

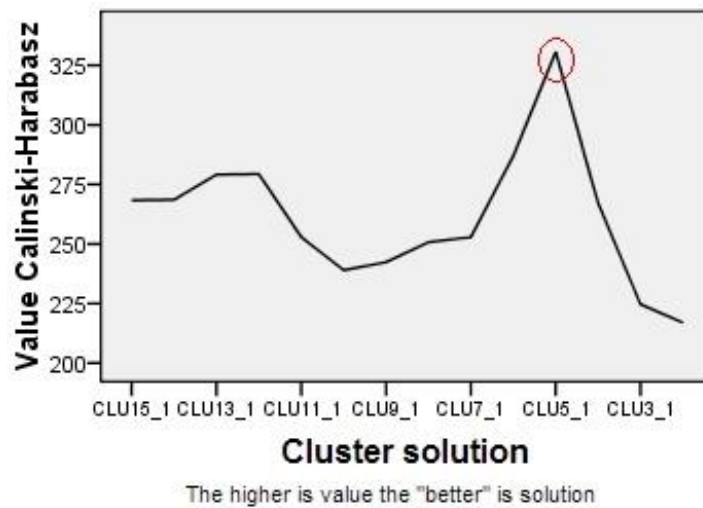


Figura 59. Ejemplo de la evolución del indicador de Calinsky-Harabasz

Otros ejemplos más recientes son los de Kolesnikov y col. (131) donde se estima el número de grupos mediante el uso modelos econométricos log-lineal del error. O Subbalakshmia y col. (132), donde usan técnicas *fuzzy* o difusas para bases de datos dinámicas. También, en Kryszczuk y Hurley (133) se propone el uso de funciones agregativas de múltiples indicadores.

En definitiva, son muchos los indicadores definidos hasta la fecha con este fin. Por este motivo le dedicaremos la siguiente sección, donde revisaremos los más utilizados y que mejor rendimiento están ofreciendo.

1.4.4.4 Segmentación espectral

La segmentación espectral es otra técnica novedosa que puede ser implementada en multitud de algoritmos de segmentación por su simplicidad y eficiencia.

En este caso, se transforma el problema de segmentación en un grafo de proximidad o similitud, $G = (V, E)$. La arista que une dos vértices cualesquiera, v_i y v_j , será ponderada con un peso no negativo, $w_{ij} > 0$, si la función de similitud entre ambos supera cierto umbral. En caso contrario, se considerará que ambos vértices no están conectados y se le asignará un peso igual a 0.

Una vez construido el grafo y su matriz de similitudes, éste es dividido o particionado de tal forma que las aristas de grupos diferentes tengan pesos muy bajos y las aristas entre puntos de un mismo grupo tengan pesos muy pequeños. Véase la Figura 60 a modo de ejemplo donde ciertas uniones entre datos no se contemplan (y tienen peso igual a 0).

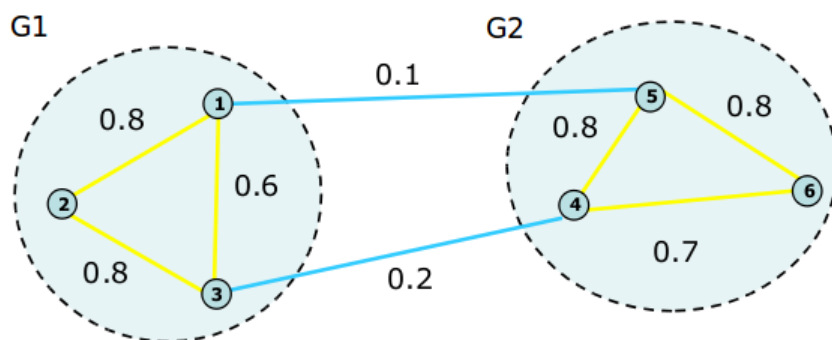


Figura 60. Grafo de proximidad o similitud

A partir del cálculo de los autovalores y la multiplicidad de los autovectores nulos de cierta matriz se pueden calcular las componentes conexas (clúster) del grafo.

Este método no es definitivo pues depende de ciertos parámetros (función de similitud y umbral considerado) pero se ha demostrado muy potente en comparación con otras técnicas clásicas. Una revisión muy sencilla de este método puede encontrarse en Von Luxburg (134).

1.4.4.5 Algoritmos evolutivos

En secciones posteriores hablaremos de los algoritmos evolutivos y se enfatizará en que es uno de los grupos de algoritmos de segmentación más prometedores hasta la fecha. Dichos algoritmos presentan unas características únicas (de potencia, robustez, velocidad, fiabilidad, flexibilidad, etc.) para cualquier problema de optimización que los convierten en una de las herramientas más adecuada para resolver varios problemas a la vez: la segmentación de los datos (usando uno o varios criterios) así como la determinación del número óptimo de clúster. Esto da origen a una rama de investigación denominada segmentación evolutiva (*Evolutionary Clustering*, en inglés).

Considerando el número de clústeres como una variable más del problema se incrementa la complejidad del problema original (al igual que ocurre cuando se incluyen varios objetivos a optimizar y se resuelve como un problema de optimización multiobjetivo). No obstante, los algoritmos evolutivos están diseñados para explorar dicho espacio (ampliado) de posibles soluciones e intensificar la búsqueda en las zonas más prometedoras en una sola ejecución del algoritmo. Esto les permite encontrar, si no la solución óptima, una solución aproximada en un tiempo antes no imaginado para un problema de tanta complejidad. Algunos ejemplos se pueden encontrar en Das y col. (134), Handl y Knowles (135), Mataka y col. (145) o Suresh y col. (137).

1.4.5 Indicadores de validez

Son muchos los indicadores de validez de la segmentación propuestos hasta la fecha. Esto es debido a su gran utilidad en multitud de casos prácticos puesto que ayudan, asignando un valor numérico a cada solución posible del problema, a determinar y medir la calidad de la partición encontrada. Estos indicadores suelen considerar el número de clústeres y diversas medidas de error o similitud para medir la validez de una solución dada. El óptimo (máximo o mínimo) de dicha función se alcanza con el número óptimo de clúster.

No obstante, no existe el indicador universal puesto que cada uno posee una serie de características que lo hacen adecuado para cierto tipo de problemas, ya sean datos

cualitativos o cuantitativos, datos binarios o no, bases de datos masivos o no, clúster compactos o alargados, etc.

No es nuestra intención hacer una revisión exhaustiva de todos los indicadores publicados hasta la fecha, más aun teniendo en mente nuestro interés en el uso de éstos para una oficina de farmacia. Recomendamos las siguientes lecturas para los que deseen profundizar en este punto: en Vendramin y col. (128) se revisan los principales indicadores de validez para problemas de segmentación. Otra revisión de éstos puede verse en Desgraupes (138) o en Das y col. (134).

Comenzamos enumerando las principales funciones distancia (de similitud o disimilitud) que pueden usarse para los indicadores y que aparecen en la Tabla 43:

| Medida | Definición | Observaciones |
|--------------------------------|---|---|
| Minkowski (Lp) | $D_{ij} = \left(\sum_{l=1}^d x_{il} - x_{jl} ^p \right)^{1/p}$ | Las variables con mayor escala tienden a dominar al resto. |
| Euclidea (L2) | $D_{ij} = \left(\sum_{l=1}^d x_{il} - x_{jl} ^2 \right)^{1/2}$ | Invariante a traslaciones y rotaciones de los datos. Las variables con mayor escala tienden a dominar al resto. |
| Tchebychev (L1) | $D_{ij} = \sum_{l=1}^d x_{il} - x_{jl} $ | Adecuado para clústers rectangulares. |
| L_{∞} | $D_{ij} = \max_{l=1, \dots, d} x_{il} - x_{jl} $ | Muy sensible a los valores extremos (outliers). |
| Mahalanobis | $D_{ij} = (x_i - x_j)^T S^{-1} (x_i - x_j)$ siendo S la matriz de covarianzas intra-grupos | Invariante a transformaciones lineales. Favorece la formación de grupos elipsoidales. |
| Coseno | $D_{ij} = \cos \alpha = \frac{x_i^T x_j}{\ x_i\ \ x_j\ }$ | Invariante a rotaciones pero no a traslaciones. |

Tabla 43. Principales funciones distancia

En la Figura 61 mostramos las distintas geometrías que presentan las circunferencias de radio según distintos valores de p .

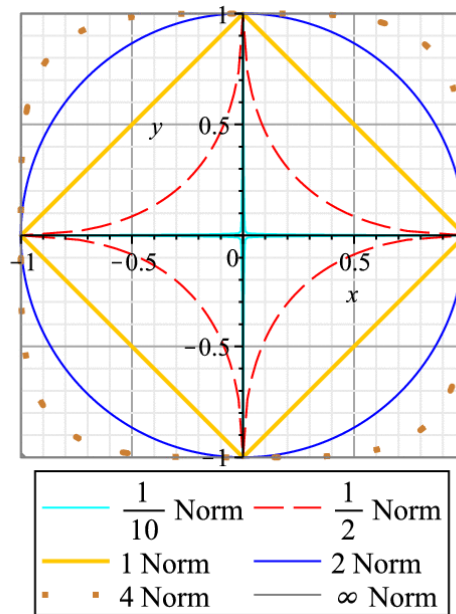


Figura 61. Distintas geometrías que presentan las circunferencias de radio según valores de p

En cambio, obsérvese la geometría de las mismas esferas pero usando la distancia de Mahalanobis en la Figura 62.

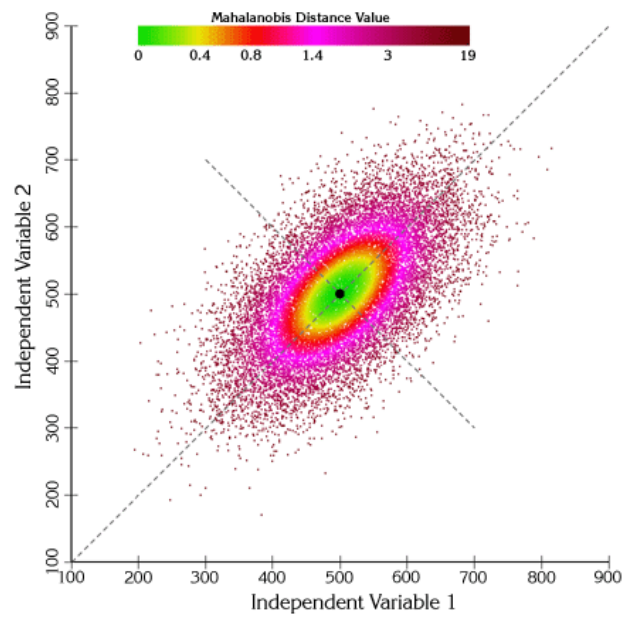


Figura 62. Geometría usando la distancia de Mahalanobis

La distancia coseno depende del ángulo que forman los puntos respecto al origen y no tanto de la distancia lineal entre ellos (Figura 63).

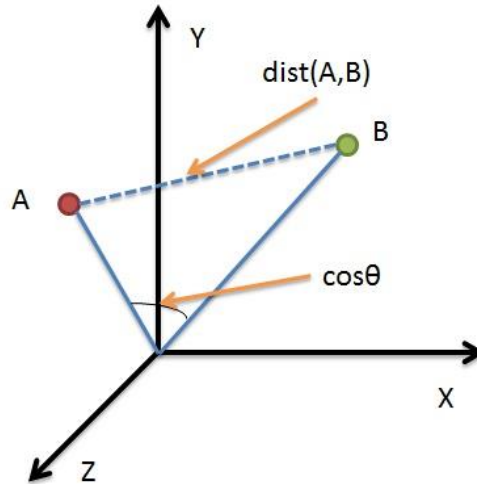


Figura 63. Distancia coseno entre dos puntos

Una vez seleccionada la función a utilizar que determina el grado de similitud entre dos puntos de la base de datos, veamos los indicadores más utilizados para medir la calidad de una partición en problemas como los que centran esta disertación.

1.4.5.1 Indicador de Calinski-Harabasz

El índice de *Calinski-Harabasz* o criterio de proporción de la varianza (130) se define como:

$$CH = \frac{\text{traza}(S_B) N_p - 1}{\text{traza}(S_W) N_p - k}$$

donde S_B y S_W son las matrices entre e intra-cluster, N_p el número de puntos y k el número de clústeres. Recordemos que la traza de una matriz se define como la suma de los elementos de su diagonal principal. El valor máximo del indicador de *Calinsky-Harabasz* determina la solución óptima teniendo en cuenta que el término $(N_p - 1)/(N_p - k)$ evita que el indicador tome valores mayores a medida que aumente el número de clústeres. Dicha solución óptima se espera obtener con grupos compactos y suficientemente separados entre sí.

En la Figura 59 ya mostramos un ejemplo de la evolución del indicador de *Calinsky-Harabasz* donde el máximo del indicador se alcanza con 5 clúster.

Respecto a la complejidad computacional del indicador, basta con tener en cuenta que tanto el cálculo de los centroides como el cálculo de las trazas tienen una complejidad de orden $O(d Np)$, donde Np es el tamaño de la base de datos y d la dimensión o número de variables consideradas. Por tanto, esta es también la complejidad global del indicador de *Calinski-Habarasz*.

1.4.5.2 Indicador de Davies-Bouldin

El indicador de *Davies-Bouldin* (139) está, en cierta manera, relacionado con el de *Calinski-Harabasz* puesto que está basado en el cálculo de la proporción de distancias intra y entre-grupos. Más concretamente,

$$DB = \frac{1}{k} \sum_{l=1}^k D_l$$

donde $D_l = \max_{l \neq m} D_{l,m}$ y $D_{l,m}$ agrega las dispersiones intra y entre los grupos l y m como sigue:

$$D_{l,m} = \frac{\bar{d}_l + \bar{d}_m}{d_{l,m}}$$

donde \bar{d}_l y \bar{d}_m son las distancias medias intra-grupo para el l -ésimo y m -ésimo clúster, respectivamente, y $d_{l,m}$ representa la distancia entre los grupos l y m . En esta ocasión, puesto que D_l representa el peor de los casos respecto al l -ésimo clúster y dicho valor se pretende minimizar, cuanto menor es el indicador DB mejor será la partición encontrada.

Respecto a la complejidad computacional, se puede comprobar que es de orden $O(d(Np+k^2))$. Como generalmente el valor de k es mucho menor que Np , este orden se reduce a $O(dNp)$.

1.4.5.3 Indicador de Dunn

Una de las familias de indicadores más popular surge del indicador de Dunn (140), también basado en las propiedades geométricas de los grupos. Se define como:

$$DN = \min_{\substack{p,q=1,\dots,k \\ p \neq q}} \left\{ \frac{\delta_{p,q}}{\max_{l=1,\dots,k} \Delta l} \right\}$$

donde Δl es el diámetro del l -ésimo clúster y $\delta_{p,q}$ es la distancia entre los clústeres p y q . Aunque otras definiciones son también válidas, se define el diámetro de un clúster como la máxima distancia entre dos puntos cualesquiera del clúster y la distancia entre los clústeres p y q como la distancia mínima entre un punto cualquiera del clúster p y otro del clúster q . Por tanto, el indicador de Dunn se desea maximizar y cuando mayor sea su valor se entiende que la partición es mejor.

Estas definiciones están claramente relacionadas con los conceptos clásicos de distancia intra y entre clúster tantas veces comentada. Así, usando el indicador de Dunn se espera también obtener particiones con clústeres compactos y separados entre sí. No obstante, las definiciones anteriores de diámetro y distancia entre clústeres han sido criticadas pues solo usan información de unos pocos puntos de los clústeres y por tanto resulta muy sensible a valores extremos o “ruido” en los datos. En la siguiente sección, veremos otras posibles definiciones.

Con respecto a la complejidad computacional se puede comprobar que es de orden $O(dNp^2+k^2)$. Como generalmente el valor de k es mucho menor que Np , este orden se reduce a $O(dNp^2)$. Obtenemos aquí otra de las más importantes desventajas del Indicador de Dunn pues su coste computacional es mucho mayor que los costes de los otros dos indicadores vistos hasta el momento.

1.4.5.4 Variantes del indicador de Dunn

Como se ha comentado previamente, en el Indicador de Dunn, las definiciones del diámetro del l -ésimo clúster, Δl , y de la distancia entre los clústeres p y q , $\delta_{p,q}$, no son las únicas posibles. Es más, las originalmente propuestas fueron criticadas pues no usaban toda la información disponible de los elementos de cada clúster sino únicamente los de los elementos “frontera”. Por este motivo, se pueden encontrar en la literatura otras cinco medidas para la distancia entre dos clústeres, $\delta_{p,q}$, así como otras dos posibles funciones para medir la varianza intra-clúster. Todas ellas, combinadas con las dadas anteriormente, darían un total de 18 variantes del indicador de Dunn (17 si no tenemos en cuenta la original).

Las cinco medidas para la distancia entre dos clústeres, $\delta_{p,q}$, son:

$$1. \delta_{p,q} = \max_{x_i \in Cp, x_j \in Cq} dist(x_i, x_j)$$

$$2. \delta_{p,q} = \frac{1}{NpNq} \sum_{x_i \in Cp} \sum_{x_j \in Cq} dist(x_i, x_j)$$

$$3. \delta_{p,q} = dist(\text{centroide } Cp, \text{centroide } Cq)$$

$$4. \delta_{p,q} = \frac{1}{Np + Nq} \left(\sum_{x_i \in Cp} dist(x_i, \text{centroide } Cq) + \sum_{x_j \in Cq} dist(x_j, \text{centroide } Cp) \right)$$

$$5. \delta_{p,q} = \max \left\{ \max_{x_i \in Cp} \min_{x_j \in Cq} dist(x_i, x_j), \max_{x_j \in Cq} \min_{x_i \in Cp} dist(x_i, x_j) \right\}$$

Mientras que las otras definiciones para la variabilidad intra-clúster son:

$$1. \Delta l = \frac{1}{Nl(Nl-1)} \sum_{x_i \neq x_j \in Cl} dist(x_i, x_j)$$

$$2. \Delta l = \frac{2}{Nl} \sum_{x_i \in Cl} dist(x_i, \text{centroide } Cl)$$

Para la complejidad computacional hay que tener en cuenta la complejidad de cada una de las funciones que se combinen pero, salvo excepciones, la complejidad de las variantes es la misma que la del indicador de Dunn original.

1.4.5.5 Indicador CS

El indicador CS (141) vuelve a considerar las nociones de distancia intra y entre clúster usando en esta ocasión los centroides como puntos representativos de los nodos.

$$CS = \frac{\sum_{i=1}^k \frac{1}{N_i} \sum_{x_i \in C_i} \max_{x_j \in C_j} dist(x_i, x_j)}{\sum_{i=1}^k \min_{j \neq i} dist(centroide C_i, centroide C_j)}$$

Al igual que los indicadores CH, DB y DN, el indicador CS compara la suma de la dispersión intra-clúster con la inter-clúster. Es más eficiente que otros indicadores con clúster de distinta densidad aunque su coste computacional es más elevado.

Como se comentó al inicio de esta sección, son muchos los indicadores propuestos, cada uno con sus características propias que lo hacen adecuados para unos problemas pero no para otros. De nuevo remitimos al lector interesado en estas medidas a los trabajos de Vendramin y col. (128) o Desgraupes (138).

1.4.6 Segmentación multiobjetivo

Nos gustaría resaltar una vez más para comenzar esta sección, que no existe una medida universal que considere todos los aspectos (densidad, compacidad, geometría...) de las posibles particiones. Más aún, todas las medidas introducidas en la sección anterior, CH, DB, DN y CS, presentan un grave problema con las unidades de medida (o magnitud) de las varianzas intra e inter-clúster y no son habitualmente normalizadas para poder compararlas adecuadamente. Más concretamente, en el indicador CS anterior las distancias intra ($dist(x_i, x_j)$) que se promedian en el numerador pueden ser de menor orden que las distancias entre centroides del denominador. Basta pensar en un problema con clústeres pequeños pero muy separados entre sí. Cuando esto ocurre y una de las varianzas domina a la otra, el resultado que se obtiene no tiene por qué ser óptimo.

Esta problemática es la que ha dado pie a una nueva línea de investigación denominada *segmentación multiobjetivo*.

En esta nueva línea, el problema de la segmentación o particionado de los datos no se realiza bajo un único punto de vista sino desde varios. Así, volviendo a considerar las medidas clásicas de variabilidad intra e inter-clúster, podríamos calcular la mejor partición que minimice la intra-clúster así como la mejor partición que maximice la inter-clúster. Hay que resaltar que ambas soluciones no suelen coincidir y que entre ambas soluciones “extremas” puede existir todo un conjunto de soluciones alternativas o de compromiso entre ambos objetivos.

En la Figura 64, tomada de Handl y Knowles (142), se resume muy acertadamente este punto.

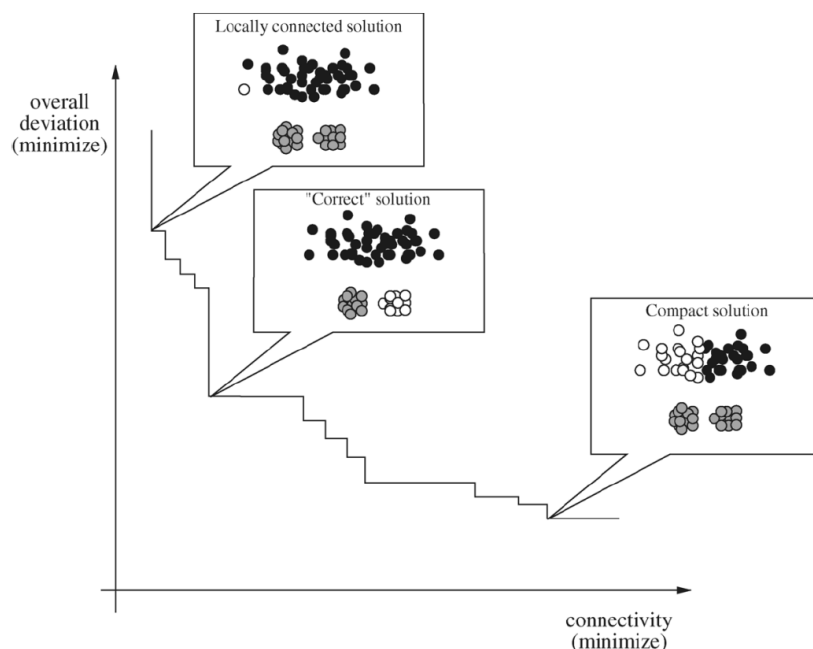


Figura 64. Particionado multiobjetivo

En esta gráfica se aprecian tres posibles soluciones válidas, según se pretenda primar la conectividad frente a la compacidad de los grupos.

Otro de los puntos a resaltar dentro de la Segmentación Multiobjetivo es que puede ayudar a solventar otro de los grandes problemas ya descrito anteriormente, el desconocimiento a priori del número de clústeres. Así, se podría considerar un tercer objetivo a optimizar.

Formalmente, si consideramos n funciones objetivo (supongamos a minimizar), f_1, f_2, \dots, f_n , que para la partición de los datos, C , devuelven los valores $f_1(C), f_2(C), \dots, f_n(C)$; se trata de determinar el conjunto de todas las posibles particiones no mejorables en un objetivo sin empeorar los otros. Para ello, dadas dos particiones $C1$ y $C2$, decimos que $C1$ domina a $C2$ si y solo si

$$f_i(C1) \leq f_i(C2) \quad \text{para todo } i, \text{ y}$$

$$f_{i^*}(C1) < f_{i^*}(C2) \text{ para un objetivo } i^*$$

El conjunto de soluciones Pareto óptimas se define como el conjunto de todas las particiones no dominadas por ninguna otra, mientras que se denomina Frente de Pareto a la imagen de las soluciones Pareto óptimas. En la Figura 65 se expone un ejemplo bi-objetivo.

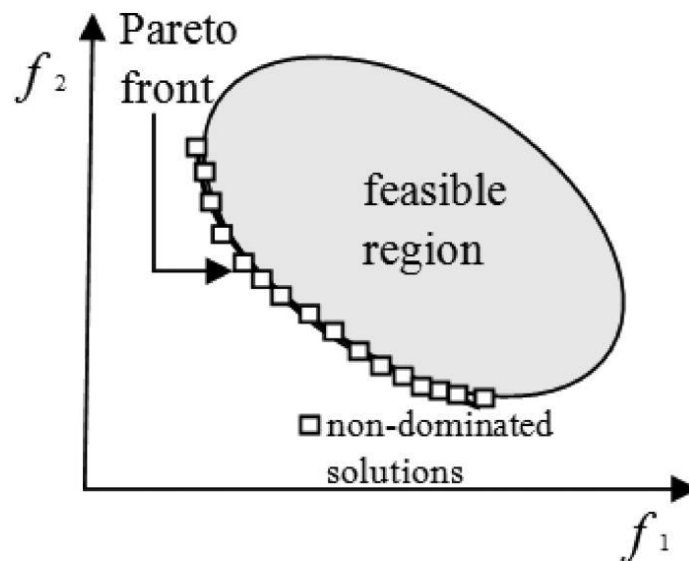


Figura 65. Soluciones Pareto y Frente de Pareto

Este enfoque parece el ideal para este y otros tipos de problemas de optimización pero no todo son ventajas. El éxito de este enfoque depende en gran medida de la capacidad del algoritmo de segmentación utilizado para encontrar las soluciones Pareto óptimas.

Tanto el enfoque multiobjetivo como el desarrollo de un algoritmo de segmentación multiobjetivo competitivo no es el objetivo de este trabajo aunque se abordará en un futuro próximo.

Finalizamos esta sección con algunas de las referencias más destacables en este campo. En Handl y Knowles (142), se propone un algoritmo de segmentación evolutivo con un mecanismo automático para determinar el número de clústeres. De manera similar pero usando otro tipo de algoritmo evolutivo está el trabajo de Saha y Bandyopadhyay (143). En Law y col. (144) se propone un algoritmo heurístico en dos fases considerando múltiples funciones objetivo. O los trabajos orientados a aplicaciones reales como Handl y Knowles (145), Suresh y col. (137), o Saha y Bandyopadhyay (146).

1.4.7 Algoritmos de segmentación

Los algoritmos de segmentación, que también encontramos en la literatura denominados como de agrupamiento, *clustering* o de aprendizaje no supervisado, permiten dividir un conjunto de objetos, representados por un conjunto de variables (características o atributos) en grupos de objetos similares llamados clústeres. De forma que los objetos pertenecientes a un mismo grupo sean más similares entre sí que respecto a los objetos de grupos diferentes (158) (Figura 66). Estos clústeres formados podrían posteriormente utilizarse como base para encontrar reglas de clasificación utilizando para ellos técnicas de aprendizaje supervisado (algoritmos de clasificación) ya que esta técnica es básicamente exploratoria.

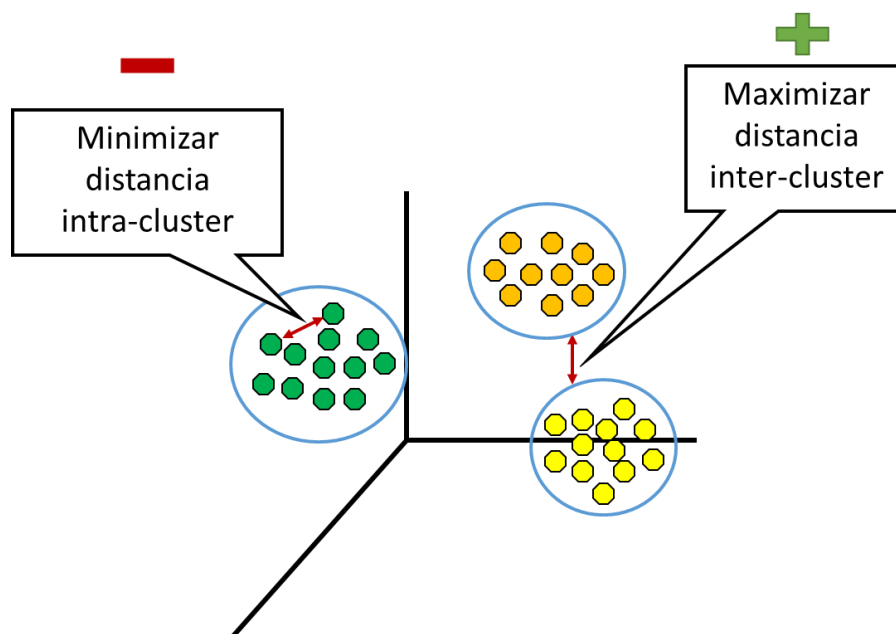


Figura 66. La formación de clústeres

Como ya se ha comentado anteriormente en esta investigación, el problema de segmentación de un conjunto de datos puede ser muy relevante para el conocimiento del comportamiento de una población; como lo pueden ser los clientes de las oficinas de farmacia.

En la Tabla 44 se ofrecen algunos conceptos básicos de segmentación o *clustering*, que aunque pueden no ser universales, nos dan una idea de una manera sistemática (134).

| Término | Definición |
|--|---|
| Patrón | Un patrón, o vector de características, es una estructura física o abstracta de objetos, los cuales serán agrupados adecuadamente por los algoritmos de <i>clustering</i> |
| Característica | Una característica o atributo es un componente individual de un patrón. Representa uno de los rasgos en base al cual los patrones se agrupan |
| Clúster | Un clúster es un conjunto bien definido de patrones similares donde los patrones de dos grupos diferentes deben ser diferentes |
| Segmentación nítida (<i>hard clustering</i>) | Un algoritmo de segmentación nítido (<i>hard</i>) es aquel que asigna cada patrón a uno y solo un clúster |
| Segmentación difusa (<i>fuzzy clustering</i>) | Un algoritmo de segmentación difusa (<i>fuzzy</i>) es aquel que asigna cada patrón a un clúster con un cierto grado de pertenencia |
| Medida de distancia | Es una métrica basada en la similitud o disimilitud de los patrones que se evalúan |

Tabla 44. Conceptos básicos sobre *clustering*

Con respecto a la formulación de un problema de segmentación, hemos optado por la utilizada por Berkhin (131), aunque ciertamente es algo bien extendido en la literatura. Siendo su consideración la siguiente:

Sea $X = \{x_1, x_2, \dots, x_N\}$ un conjunto de datos $x_i = (x_{i1}, x_{i2}, \dots, x_{id}) \in A$, también llamados patrones, objetos, casos, n-uplas, puntos, etc., donde A es un espacio de características o atributos de la población observada, $i = 1, \dots, N$, y cada componente x_{ij} ($j = 1, \dots, d$) es un atributo (variable, dimensión o componente) de tal modo que el conjunto de objetos X forma una matriz $N \times d$ que es empleada por la mayoría de los algoritmos de segmentación.

La finalidad de todo algoritmo de segmentación es asignar cada objeto a un conjunto finito de K subconjuntos o clústeres de modo que objetos similares pertenezcan al mismo clúster, mientras que los objetos de clústeres diferentes sean lo menos parecidos posible.

Del análisis de los trabajos realizados por diferentes autores sobre la división en grupos de una población, se observa que cada técnica está diseñada para realizar una clasificación de tal modo que cada grupo sea lo más homogéneo y lo más diferente de los demás como sea posible.

El resultado de cada algoritmo de segmentación dependerá de tres factores:

- El tipo de algoritmo elegido.
- El valor de los parámetros (número de clústeres, matriz de datos y valores iniciales de los clústeres).
- Las funciones distancia elegidas (similaridad / disimilaridad). Una selección representativa de estas medidas las vimos con anterioridad en la Tabla 43.

Dentro del *clustering* o segmentación de poblaciones o bases de datos, se engloban diferentes métodos, técnicas y algoritmos que se distinguen tanto en la forma de contemplar los clústeres como en la forma de encontrarlos. Una primera diferencia, surge entre aquellos métodos que parten de un número predeterminado de clústeres y los que no, dando lugar a la clasificación entre métodos de particionamiento y métodos jerárquicos.

Otra diferencia aparece en la exclusividad de pertenencia o no de un objeto a un grupo o clúster. Los algoritmos de segmentación por partición asignan cada objeto del

conjunto a uno y solo uno de los grupos, asumiendo así que los límites entre los grupos están bien definidos. A menudo, esta asunción no se corresponde con la realidad de los datos, donde los límites entre los grupos pueden ser difusos y se requiere una descripción más matizada del objeto al grupo específico al que pertenece. De aquí surge la idea del segmentación difusa (*fuzzy clustering*) como alternativa para modelar lo más real posible la casuística comentada. Concretamente el agrupamiento difuso acepta el hecho de que los grupos o clústeres no están completamente separados (Figura 67) asignando por ello a cada objeto del conjunto un grado de pertenencia, habitualmente entre 0 y 1, a los diferentes grupos.

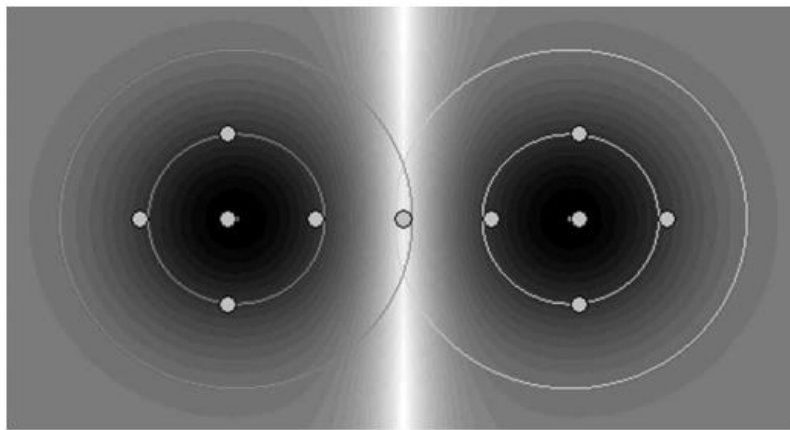


Figura 67. Agrupamiento difuso de dos grupos. Las zonas oscuras indican altos grados de pertenencia.
Basado en Sosa y col. (159)

Esta discusión, da lugar a otra clasificación diferente:

- Algoritmos de pertenencia exclusiva a clústeres.
- Algoritmos de pertenencia no exclusiva a clústeres.

Los algoritmos de pertenencia no exclusiva más característicos son los algoritmos de agrupamiento difuso (*fuzzy clustering*) (159) , en donde cada objeto tiene asignado un grado de pertenencia a todos los grupos establecidos; el grado de pertenencia es mayor cuanto más similar es el objeto al patrón o representante del clúster al que se refiere.

Además de estas clasificaciones, que se podríamos considerar clásicas, existen otras que se basan en nuevas técnicas y que dan lugar a nuevas clasificaciones de los algoritmos

de segmentación basadas en los trabajos de Berkhin (131) y Xu y Wunsch (130). En la Figura 68 se expone la clasificación ampliada.

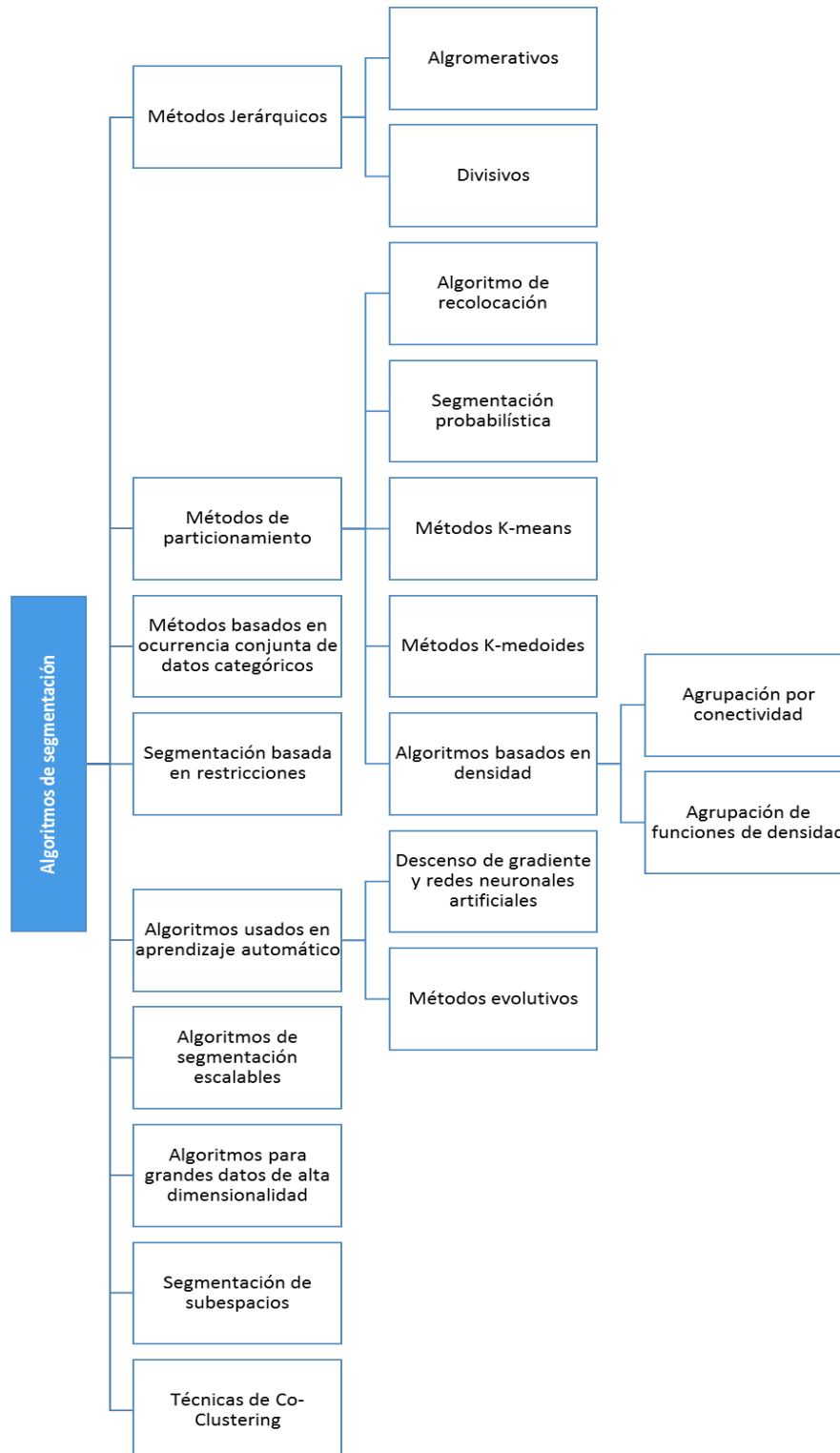


Figura 68. Clasificación de los algoritmos de segmentación

En la Tabla 45 podemos ver los algoritmos más utilizados para cada uno de estos métodos de segmentación.

| Clasificación de segmentación | Algoritmos |
|--|---|
| Métodos Jerárquicos | COBWEB, CLASSIT (ambos no se consideran ni aglomerativos ni divisivos) |
| Aglomerativos | Linkage simple, completo, promedio, mediano, centroide, método de Ward, BIRCH, CURE, ROCK, CHAMELEON, AGNES ... |
| Divisivos | DIANA, MONA |
| Métodos de particionamiento | |
| Algoritmos de recolocación | |
| Segmentación probabilística | EM, SNOB, AUTOCLASS, MCLUST |
| Métodos K-means | K-means, Forgy, ISODATA, GKA |
| Métodos K-medoides | PAM, CLARA, CLARANS |
| Algoritmos basados en densidad | |
| Agrupación por conectividad | DBSCAN, GDBSCAN, OPTICS |
| Agrupación de funciones de densidad | DENCLUE |
| Métodos basados en rejillas | CLIQUE, MAFIA, BANG, GRIDCLUST, STING, WAVECLUSTER, FC |
| Métodos basados en ocurrencia conjunta de datos categóricos | CACTUS, HMETIS, STIRR |
| Segmentación basada en restricciones | COP K-means, PCKmeans, CMWK-Means |
| Algoritmos usados en aprendizaje automático | CLTree, AMOEBA |
| Descenso de gradiente y redes neuronales artificiales | LKMA, SOM |
| Métodos evolutivos | SINICC, CGA |
| Algoritmos de segmentación escalables | DIGNET, BUBBLE |
| Algoritmos para grandes datos de alta dimensionalidad | |
| Segmentación de subespacios | CLIQUE, ENCLUS, MAFIA, OPTIGRID, PROCLUS, ORCLUS |
| Técnicas de Co-Clustering | PING-PONG, SIMPLIFYRELATION |

Tabla 45. Métodos y algoritmos de segmentación: clasificación ampliada

A continuación vamos a ver con más detalle un conjunto de algoritmos con la intención final de comparar su aplicación con la del algoritmo propuesto en esta investigación.

1.4.7.1 Algoritmos de particionamiento: *K-means* y variantes

Para este tipo de algoritmos, la segmentación o *clustering* de X es una partición de X en K grupos, segmentos o clúster $\{C_1, C_2, \dots, C_k\}$ tales que:

1. Los clústeres no están vacíos: $C_i \neq \emptyset, 1 \leq i \leq k$.
2. Los clústeres cubren todo X : $\bigcup_{i=1}^k C_i = X$ (salvo excepción de *outliers*).
3. Los clústeres son disjuntos dos a dos: $C_i \cap C_j \neq \emptyset$, si $i \neq j$ (en algún caso esta condición no se cumple).

1.4.7.1.1 Algoritmo *K-means*

El algoritmo de las K -medias, conocido como *K-means*, es uno de los algoritmos más populares y utilizados en análisis clúster o segmentación. Aunque muy útil, la primera versión del algoritmo fue propuesta por Stuart Lloyd en 1957 publicado en 1982 (161) aunque no fue computacionalmente óptima. La versión actual fue propuesta por MacQueen (162). Es un algoritmo del tipo no jerárquico y está indicado en aquellas situaciones en las que todas las variables son del tipo cuantitativo, el número K de clústeres es establecido a priori y como medida de diferencia se toma la distancia euclídea de forma habitual.

$$d(x_i, x_{i'}) = \sum_{j=1}^d (x_{ij} - x_{i'j})^2 = \|x_i - x_{i'}\|^2$$

A los centros multivariantes de los clústeres se les denominan centroides

Los pasos que sigue el algoritmo *K-means* son los siguientes:

1. Se eligen K puntos del conjunto o base de datos que se está segmentando. Estos puntos representan los centroides (medias) iniciales.
2. Se asigna cada objeto al clúster que tiene el centroide más cercano.
3. Cuando todos los objetos han sido asignados en los k clústeres, se recalculan las posiciones de los K centroides.

4. Se repiten los pasos 2 y 3 hasta que los centroides no varíen.

El centroide puede variar dependiendo de la medida de proximidad para los datos y del objetivo del *clustering*. El objetivo del *clustering* se expresa generalmente por una función objetivo que depende de la proximidad entre los puntos o la proximidad a los clústeres. Si se consideran datos cuya medida de proximidad es la distancia euclídea, para una función objetivo que mide la calidad del agrupamiento, se puede usar la suma del error cuadrado (SSE). Esto es, se calcula el error de cada punto como su distancia euclídea al centroide más cercano, y luego se calcula la suma total de los errores cuadrados. De esta forma la SSE se puede definir como:

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} d(c_i, x)^2$$

donde C_i es el i -ésimo clúster, c_i el centroide del clúster C_i , x una observación (patrón) cualquiera y K el número de clústeres.

El algoritmo *K-means* trata de minimizar la SSE. Así, el criterio es asignar las N observaciones a los K clústeres de modo que dentro de cada clúster el promedio de las diferencias de cada observación a la media del clúster, definido por los puntos del clúster, sea mínima.

El requerimiento de espacio de *K-means* es pequeño ya que solamente se almacenan los puntos y los centroides. Concretamente el espacio requerido es $O((N + K)d)$, donde N es el número de puntos y d es el número de atributos o variables. En cuanto al costo computacional, *K-means* también requiere poco tiempo, básicamente lineal $O(l * k * N * d)$, donde l es el número de iteraciones requeridas para la convergencia.

| n | 10 | 100 | 1000 | 10000 | 100000 |
|-----------------------------------|-----------|------------|-------------|--------------|---------------|
| $O(n)$ | 10 ms | 0,1 s | 1s | 10s | 100s |
| $O(n \log_2 n)$ | 33ms | 0.7s | 10s | 2 min | 28 min |
| $O(n^2)$ | 100ms | 10s | 17 min | 28 horas | 115 días |
| $O(n^3)$ | 1s | 17 min | 12 días | 31 años | 32 milenios |

Tabla 46. Impacto de la eficiencia de un algoritmo

Este algoritmo plantea las siguientes cuestiones:

- No se especifica la elección de los centroides iniciales.
 - ¿Elección aleatoria de los K primeros centroides?
 - Los resultados de la segmentación dependen de la elección inicial de los centroides.
 - ¿Ejecutar el algoritmo varias veces con diferentes centroides iniciales?
- Se asume que K es conocido a priori.
 - ¿Cómo lo elegimos?

Son variados los métodos utilizados para encontrar el número adecuado de clústeres. Algunos de ellos son (130):

- Visualización del conjunto de datos, lo que funciona bien para dos dimensiones pero generalmente nuestros conjuntos de datos son mucho más complicados.
- Construcción de índices (o reglas de paro). En este caso se utilizan índices para enfatizar el grado de compacidad intra-clúster y separación (aislamiento) inter-clúster considerando efectos tales como: el error cuadrático, propiedades geométricas o estadísticas de los datos, el número de patrones, la disimilaridad o similaridad, número de clústeres.
- Optimización de alguna función de criterio bajo el marco del modelo de mezcla de probabilidades. Usualmente se utiliza el algoritmo EM para encontrar el valor de K que maximice o minimice el criterio definido como óptimo.
- Criterio de Información de Akaike (AIC).
- Criterio de Inferencia Bayesiana.
- Otros métodos heurísticos basados en una variedad de técnicas y teorías.

Las respuestas posibles a estas preguntas se basan en la experiencia de su uso del algoritmo y del tipo de datos que se están segmentando. No hay reglas preestablecidas para ello, en cualquier caso, siempre se puede ejecutar el algoritmo con bases de datos

maestras, como la que hemos visto en el apartado anterior, y estudiar el comportamiento del algoritmo.

Con respecto a la elección de K primeros centroides, varios autores proponen usar diferentes alternativas: las primeras K observaciones, elegir aleatoriamente K observaciones o tomar cualquier partición al azar en K clústeres y calculando sus centroides.

Las características generales de este algoritmo son las siguientes:

- No se satisface el criterio de optimización globalmente, solo produce un óptimo local.
- El algoritmo de *K-means* es computacionalmente rápido.
- Puede trabajar bien con valores perdidos (*missing values*).
- Es sensible a “*outliers*” (valores anómalos).

La mayoría de los algoritmos no jerárquicos difieren entre sí según:

- El procedimiento para obtener los centroides iniciales.
- La regla que se usa para reasignar las observaciones.

1.4.7.1.2 Algoritmo PAM (Partitioning Around Medoids)

Este algoritmo fue desarrollado por Kaufman y Rousseeuw (158) para encontrar K clústeres como alternativa al algoritmo *K-means*. Para ello determina un objeto representativo para cada clúster llamado medoide. Este objeto representativo es el que se encuentra localizado más al centro dentro del clúster. Una vez que los medoides han sido seleccionados, cada objeto no seleccionado es agrupado con el medoide al cual es más similar (como ya hemos visto, *K-means* utiliza los promedios del grupo como representantes de los clústeres).

De forma más concreta, si O_h es un objeto no seleccionado, y O_i es un medoide (objeto seleccionado), se dice que O_h pertenece al clúster representado por O_i si $d(O_h, O_i) = \min_{O_e} d(O_h, O_e)$; donde \min_{O_e} denota el mínimo sobre todos los medoides O_e , y $d(O_h, O_i)$

denota la disimilaridad o distancia entre los objetos O_h y O_i . Todos los valores de disimilaridad son dados como datos de entrada al PAM.

Para encontrar los K medoides, el algoritmo PAM comienza con una selección arbitraria de K objetos. En cada iteración, un intercambio entre un objeto seleccionado O_i y un objeto no seleccionado O_h es realizado, si y solo sí, el intercambio resulta en un incremento de la calidad del agrupamiento (*clustering*). Concretamente, para calcular el efecto del intercambio entre O_h y O_i , el algoritmo PAM calcula el coste C_{jih} para todos los objetos no seleccionados O_j . El costo total de reemplazar O_h por O_i se denota por $TC_{hi} = \sum_j C_{jih}$

El cálculo del coste C_{jih} presenta cuatro casos:

- Caso 1: sea O_j un objeto perteneciente al clúster representado por O_i , además, asumiendo que O_j es más similar a $O_{j,2}$ que a O_h , $d(O_j, O_h) \geq d(O_j, O_{j,2})$ donde $O_{j,2}$ es el segundo medoide más similar a O_j . Entonces, si O_i es reemplazado por O_h como un medoide, O_j pertenecería al clúster representado por $O_{j,2}$. De esta manera el costo del intercambio es dado por:

$$C_{jih} = d(O_j, O_{j,2}) - d(O_j, O_i)$$

- Caso 2: O_j pertenece actualmente al cluster representando por O_i . Pero esta vez, O_j es menos similar a $O_{j,2}$ que a O_h , $d(O_j, O_h) \geq d(O_j, O_{j,2})$. Entonces, si O_i es reemplazado por O_h , O_j pertenecería al clúster representado por O_h . De esta manera el costo es dado por:

$$C_{jih} = d(O_j, O_h) - d(O_j, O_i)$$

Caso 3: suponiendo que O_j pertenece actualmente a un cluster diferente al representado por O_i y que $O_{j,2}$ es el objeto representativo de ese clúster y O_j es más similar a $O_{j,2}$ que a O_h . Entonces aún si O_i es reemplazado por O_h , j permanecería en el clúster representado por $O_{j,2}$. De esta manera el costo es:

$$C_{jih} = 0$$

Caso 4: suponiendo que O_j pertenece actualmente al clúster representado por $O_{j,2}$, donde O_j es menos similar a $O_{j,2}$ que a O_h . Entonces reemplazando O_i por O_h causaría que $O_{j,i}$ saltara al clúster representado por O_h desde el clúster $O_{j,2}$. De esta manera el coste es dado por:

$$C_{jih} = d(O_j, O_h) - d(O_j, O_{j,2})$$

Algoritmo:

1. Seleccionar aleatoriamente k objetos representativos.
2. Calcular TC_{hi} para todos los pares de objetos O_h, O_i donde O_i está actualmente seleccionado, y O_h no lo está.
3. Seleccionar el par O_h, O_i el cual corresponde a $\min_{O_h, O_i}(TC_{hi})$. Si el mínimo TC_{hi} es negativo, reemplazar O_i por O_h , y regresar al paso 2.
4. En caso contrario, para cada objeto no seleccionado, encontrar el objeto representativo más similar.
5. Fin

Al igual que *K-means*, el algoritmo de PAM necesita una configuración (partición) inicial y un número pre-especificado de clústeres. En general PAM es más robusto que *K-means* (es menos sensible al “ruido” y *outliers*) y requiere como argumento de entrada solamente la matriz de disimilaridades entre observaciones y no los datos originales. Como contrapartida es más intensivo computacionalmente, debido principalmente a la búsqueda de medoides (165). De hecho, la complejidad total del algoritmo PAM es de $O(K(N-K)^2)$ y para grandes valores de N el costo se torna muy alto (164).

1.4.7.1.3 Algoritmo CLARA (Clustering LARge Applications)

El algoritmo PAM funciona bien en bases de datos pequeñas pero es lento en las grandes bases de datos esto originó el desarrollo del algoritmo CLARA (*Clustering Large Applications*). CLARA crea múltiples muestras de los datos y entonces aplica el algoritmo PAM a la muestra.

La diferencia entre PAM y CLARA es que CLARA se basa en muestreo. Solo una pequeña parte de todos los datos es seleccionada como representativa de los datos y los

medoides son escogidos de la muestra usando el algoritmo PAM. Por lo tanto, este nuevo algoritmo se basa en la presunción de que si la muestra es seleccionada de manera aleatoria, entonces será representativa del conjunto total de datos, y los objetos representativos escogidos (medoides), serán similares tal y como si hubieran sido escogidos del conjunto total de datos. A continuación se presenta el algoritmo CLARA con n muestras de tamaño t :

Algoritmo:

-
1. Desde $i = 1$ hasta n , repetir los siguientes pasos.
 2. Seleccionar una muestra aleatoria de t objetos del conjunto total de datos y utilizar el algoritmo PAM para encontrar K medoides de la muestra.
 3. Para cada objeto O_h del conjunto total de datos, determinar cuál de los K medoides es el más similar a él.
 4. Calcular la disimilaridad promedio del agrupamiento obtenido en el paso anterior. Si este valor es menor al mínimo actual, usar este valor como el mínimo actual y retener los K medoides encontrados en el paso (2) como el mejor conjunto de medoides obtenidos.
 5. Volver al paso 1 para comenzar con una nueva iteración.
-

1.4.7.1.4 Algoritmo CLARANS (Clustering Large Applications based upon RANdomized Search)

CLARANS, que surge como modificación del algoritmo CLARA, fue propuesto por Ng y Han en 1994 (167).

El algoritmo CLARANS es una mezcla de los algoritmos PAM y CLARA y surge con el objetivo de mejorar la calidad y la escalabilidad de CLARA. Las búsquedas las realiza sobre un subconjunto del conjunto total de datos y no se limita a ninguna muestra. Mientras CLARA tiene una muestra fija en cada etapa de la búsqueda, CLARANS forma una muestra aleatoria en cada etapa de la búsqueda.

En el algoritmo CLARANS, todas las posibles soluciones candidatas son vistas como un grafo $G_{n,K}$ en el que cada nodo representa una solución candidata (168). Dos soluciones vecinas o nodos de este grafo difieren entre sí por un único medoide de K .

Por lo tanto, cada solución puede tener $K \cdot (n-K)$ vecinos, donde n es el tamaño del subconjunto muestreado del conjunto de datos original. Puesto que cada nodo representa una solución con medoides de K , cada nodo representa también un agrupamiento. Por ello, a cada nodo se le puede asignar un coste que se determina por la disimilitud total entre cada elemento no-medoide y el medoide de su clúster. Luego si los medoides x_{mi} y x_{ni} son los únicos medoides diferentes en las soluciones vecinas S_1 y S_2 , la diferencia de coste entre S_1 y S_2 es obtenida por $TC_{m_i n_i}$.

Algoritmo:

-
1. Dar como datos de entrada los parámetros *numlocal* y *maxneighbor*.

Inicializar i ($i=1$) y *mincost* a un número mayor.
 2. Establecer *current* a un nodo arbitrario en $G_{n,k}$.
 3. Establecer $j = 1$.
 4. Considerar un vecino aleatorio S de *current*, y con la expresión $TC_{hi} = \sum_j C_{jih}$ calcular el coste diferencial de los 2 nodos.
 5. Si S tiene un coste menor, establecer *current* a S , e ir al paso 3.
 6. De lo contrario, incrementar j en 1. Si $j \leq \text{maxneighbor}$ ir al paso 4.
 7. De lo contrario, cuando $j > \text{maxneighbor}$, comparar el coste de *current* con *mincost*. Si éste es menor a *mincost*, establecer *mincost* al coste de *current*, y establecer *bestnode* a *current*.
 8. Incrementar i en 1. Si $i > \text{numlocal}$, el resultado es *bestnode* y terminar. De lo contrario, ir al paso 2.
-

1.4.7.2 Algoritmos jerárquicos: aglomerativos y divisivos

Los resultados de aplicar las técnicas de particionamiento vistas en el anterior apartado, dependen de la elección del número de clústeres y de una configuración inicial. En cambio, los métodos de *clustering* jerárquico no requieren estas especificaciones.

Los algoritmos jerárquicos son algoritmos secuenciales en los que cada etapa del algoritmo consiste en unir o separar grupos. En este sentido, el algoritmo jerárquico de segmentación permite establecer una jerarquía de grupos dentro del conjunto de datos, de tal manera que cada grupo se encuentra anidado dentro de otro más grande.

Estos algoritmos intentan construir una partición de X con una estructura anidada del tipo árbol (Figura 69) también conocido como dendrograma (Figura 69).

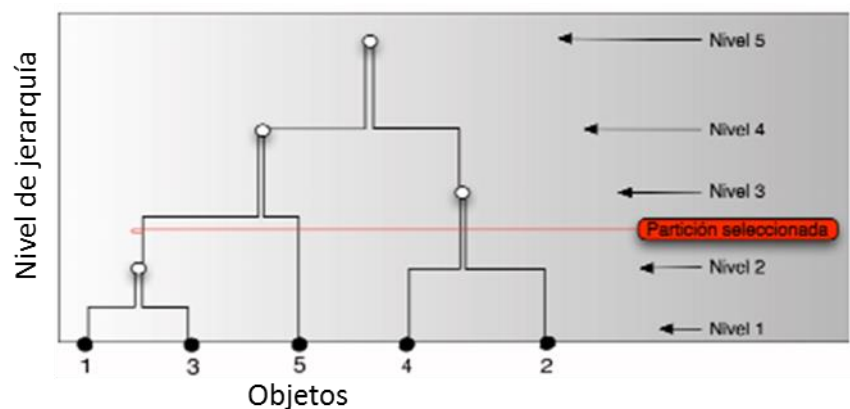


Figura 69. Ejemplo dendrograma

Los algoritmos jerárquicos se dividen en aglomerativos y divisivos (161) y (158) según el orden en que se vayan formando los grupos. Sus características principales son las siguientes:

- Métodos jerárquicos aglomerativos (*bottom-up*): en el primer paso, cada unidad forma un grupo. En cada etapa del algoritmo se unen dos grupos, y así sucesivamente hasta que se llega a un único grupo formado por todos los objetos o patrones.
- Métodos jerárquicos divisivos (*top-down*): en este caso el algoritmo comienza con un único grupo que contiene a todas las unidades. En cada etapa se divide

un grupo en dos, hasta que se llega a tantos grupos como objetos individuales había en el conjunto.

En cuanto a las ventajas y desventajas de los algoritmos jerárquicos destacan las siguientes (131):

- Ventajas.
 - Flexibilidad en cuanto al nivel de granularidad.
 - Facilidad de manejo de cualquier forma de similitud o distancia.
 - En consecuencia, la aplicabilidad a cualquier tipo de atributos.
- Desventajas.
 - La ambigüedad de los criterios de parada.
 - El hecho de que la mayoría de los algoritmos jerárquicos no vuelven a visitar los grupos una vez contruidos con el propósito de su mejora.

Métodos jerárquicos aglomerativos

Estos algoritmos presentan los siguientes pasos de forma general:

1. Agrupamiento o segmentación inicial ($t=0$): $S_0 \equiv \{C_i = \{x_i\} \text{ donde } i = 1, \dots, N\}$
2. Realizar nueva iteración: $t \rightarrow t + 1$
3. Calcular las distancias $d(C_r, C_s)$ entre todos los clústeres posibles C_r, C_s del agrupamiento anterior S_{t-1} seleccionando la pareja de clústeres C_i, C_j que presentan la mínima distancia entre ellos.
4. Se define un nuevo grupo $C_q = C_i \cup C_j$. El nuevo agrupamiento resultante es $S_t = S_{t-1} \cup C_q$
5. Si ya se dispone de un único clúster, STOP. En otro caso, volver al paso 2.

En cada iteración solamente será necesario calcular las distancias entre el nuevo clúster formado C_q y el resto. La variedad de posibilidades para calcular dicha distancia, es la que provee diferentes modalidades de implementación del algoritmo. Entre los métodos más utilizados para calcular las distancias entre clústeres, conocido como *linkage* en la literatura de *clustering*, destacan los siguientes:

- Actualización de centroides (ponderados por el número de objetos que contienen).
- Medidas basadas en la fórmula de Lance y Williams (162):
 - Algoritmo de enlace simple:

$$d(C_q, C_s) = \min(d(C_i, C_s), d(C_j, C_s)) \text{ (Linkage simple)}$$

- Algoritmo de enlace completo:

$$d(C_q, C_s) = \max(d(C_i, C_s), d(C_j, C_s)) \text{ (Linkage completo)}$$

- Algoritmo de promedio no ponderado:

$$d(C_q, C_s) = \frac{1}{2} (d(C_i, C_s) + d(C_j, C_s))$$

- Algoritmo de promedio ponderado:

$$d(C_q, C_s) = \frac{1}{n_i + n_j} (n_i \cdot d(C_i, C_s) + n_j \cdot d(C_j, C_s))$$

- Algoritmo de centroide no ponderado:

$$d(C_q, C_s) = \frac{1}{2} (d(C_i, C_s) + d(C_j, C_s)) - \frac{1}{4} d(C_i, C_j)$$

- Algoritmo de centroide ponderado:

$$d(C_q, C_s) = \frac{n_i}{n_i + n_j} d(C_i, C_s) + \frac{n_j}{n_i + n_j} d(C_j, C_s) - \frac{n_i \cdot n_j}{n_i + n_j} d(C_i, C_j)$$

La Figura 70 ilustra las distancias comentadas entre clústeres o conglomerados.

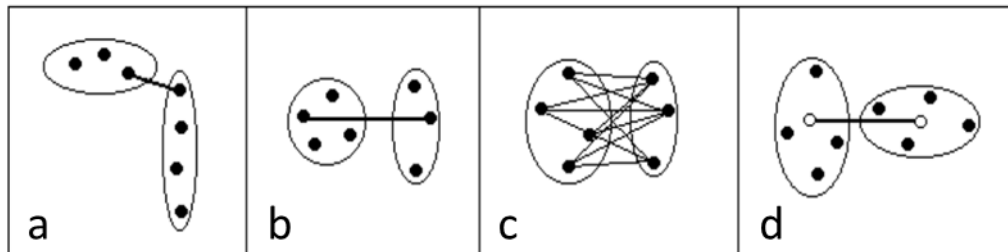


Figura 70. Distancias entre clústeres: a) Vecino más cercano b) más lejano c) promedio d) centroide o centro de gravedad

Métodos jerárquicos divisivos

Como ya hemos visto antes, los algoritmos jerárquicos divisivos parten de un único clúster con todas las observaciones y recursivamente dividen uno de los clústeres existentes en dos clústeres “hijos” hasta obtener tantos grupos como observaciones. Dividir un clúster es computacionalmente más costoso que unir dos, dado que no solo se debe encontrar el clúster a ser dividido sino que también las observaciones que formarán los dos nuevos grupos deben ser identificadas. Por este motivo, los métodos divisivos son menos utilizados en la práctica (Varmuza y Filmoser, 2009).

A continuación, vamos a mostrar diferentes variantes de algoritmos jerárquicos.

1.4.7.2.1 BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)

Este algoritmo fue propuesto por Zhang y col. (1996) y es apropiado para bases de datos de grandes dimensiones.

Almacena para cada clúster un triplete de datos que contiene:

- El número de objetos que pertenecen al clúster.
- La suma de todos los valores de los atributos de todos los objetos pertenecientes al clúster.
- La suma de los cuadrados de los atributos de los objetos que pertenecen al clúster.

Con esta información se construye un árbol de grupos llamado CF-tree (*Cluster Features tree*). En cada nodo se indica el número de grupos que pertenecen a esa ramificación y cuáles son sus características.

El procedimiento del algoritmo BIRCH es el siguiente:

1. Generar un CF-tree inicial, leyendo los datos y asignándolos a una rama o a otra. Si la distancia entre un objeto nuevo y los anteriores se hace mayor que cierto parámetro T , se crea una rama nueva.

2. Revisar el árbol creado para ver si es demasiado grande, y moldearlo modificando el valor del parámetro T . Si el valor de este parámetro se aumenta, las ramas del árbol se juntan al no haber distinción de grupos.
3. Aplicar algún procedimiento de *clustering*, como el *K-means*, sobre la información contenida en los nodos de cada nivel.
4. Redistribuir los datos según los centroides descubiertos en el paso anterior, logrando un mayor refinamiento en el agrupamiento.

Las principales desventajas del algoritmo *BIRCH* son su secuencialidad, lo cual puede provocar asignación a distintos clústeres de objetos replicados, colocados en distintos lugares de la secuencia, y la fuerte dependencia del parámetro T . De forma que una mala elección de este valor puede generar la creación de falsas agrupaciones, o ramificaciones duplicadas, o la asignación de objetos a un mismo nodo, cuando deberán estar en nodos distintos.

1.4.7.2.2 CURE (Clustering Using REpresentatives)

Se trata de una extensión del algoritmo *K-means* para clústeres de formas y tamaños arbitrarios ya que *K-means* asume que los clústeres están normalmente distribuidos en cada dimensión (ejes fijos). Lo que aporta CURE como solución, es usar la distancia euclídea y un conjunto de puntos representativos para representar a cada clúster de forma que este algoritmo permite que los clústeres adopten cualquier forma.

Los representantes de cada clúster se construyen a partir de una muestra de puntos del clúster, tan dispersa como sea posible. A partir de la muestra, se crean los representantes moviendo los puntos de la muestra hacia el centroide del clúster (por ejemplo, un 25%).

El algoritmo CURE presenta dos fases:

- Fase 1.
 - Se selecciona una muestra del conjunto de datos que quepa en memoria principal.

- Se agrupan esos puntos utilizando un algoritmo jerárquico (combinando los puntos/clústeres más cercanos).
- Se selecciona un conjunto de representantes para cada uno de los clústeres obtenidos.
- Fase 2.
 - Se recorre el conjunto completo de datos y cada punto se asigna al clúster más cercano (se busca el representante más cercano al punto y el punto se asigna al clúster del representante).

Algoritmo:

- Para cada clúster c (o cada punto de entrada):
 - $c.mean$ = media aritmética de los puntos del clúster.
 - $c.rep$ = conjunto de r puntos representativos del clúster (inicialmente $r=1$ ya que cada clúster tiene un solo dato en dicho instante).
 - $c.closest$ = el clúster más cercano a c .
 - Todos los puntos de entrada son insertados en un árbol k-d, identificado por T .
 - Tratar cada punto de entrada como clúster separado. Calcular $c.closest$ para cada c y luego insertar cada clúster en un conjunto Q (clústeres dispuestos en orden de distancia creciente entre c y $c.closest$).
 - Mientras $tamaño(Q) > k$
 - Retirar el elemento superior de Q , es decir c , y unirlo con su clúster más cercano $c.closest$ (es decir, v) y calcular los puntos representativos para el nuevo clúster formado w .
 - Retirar c y v de T y Q .
 - Para todos los clústeres x en Q ,
 - Actualizar $x.closest$ y reubicar x
 - Insertar w en Q
-

1.4.7.2.3 ROCK (RObust Clustering using linKs)

Se trata de un algoritmo de *clustering* jerárquico para conjuntos de datos con atributos booleanos y categóricos. En este algoritmo, un par de puntos se consideran vecinos si su similitud está por encima de un umbral preestablecido. La medida de similitud usada por este algoritmo es el “número de vecinos compartidos”.

Algoritmo:

-
1. Obtener una muestra del conjunto de datos.
 2. Calcular el valor del enlace entre cada par de puntos (transformar similitudes originales, calculadas usando el coeficiente de Jaccard, en similitudes que reflejen el número de vecinos compartidos).
 3. Usar un algoritmo aglomerativo utilizando el número de vecinos compartidos como medida de similitud.
-

1.4.7.2.4 CHAMELEON

Este algoritmo, desarrollado por Karypis y col. (163) presenta dos fases principales. En la primera fase construye el grafo de los K vecinos más cercanos y usa un algoritmo de particionamiento de grafo para agrupar los puntos en subgrupos.

En la segunda fase, usa un algoritmo jerárquico aglomerativo para encontrar los clústeres genuinos combinando repetidamente estos subgrupos. En esta segunda fase determina el par de subgrupos más similares tomando en cuenta su interconectividad y cercanía (éstas expresan las características internas de los subgrupos). El modelo no es estático, sino que es capaz de adaptarse a las características internas de los subgrupos según estos van cambiando.

Su pseudocódigo es el siguiente:

-
1. Obtener el grafo de los K vecinos más cercanos.
 2. Dividir el grafo utilizando un algoritmo de particionamiento de grafos multinivel.

3. (REPETIR) Anidar los dos clústeres que mejor mantengan su propia similitud respecto a la interconectividad y cercanía relativas (HASTA que no se puedan anidar más clústeres).
-

1.4.7.2.5 Algoritmo de segmentación bietápico

Aunque este algoritmo no es estrictamente un algoritmo jerárquico, lo hemos incluido en esta sección por utilizar en una de sus fases un algoritmo jerárquico para la obtención de clústeres. El hecho de comentar su procedimiento obedece a que es un algoritmo utilizado por el software comercial *IBM SPSS 22* y que utilizaremos como un algoritmo con el que comparar el propuesto en esta investigación.

El algoritmo de segmentación bietápico es un método de agrupación en clústeres de dos pasos. El primer paso consiste en hacer una única pasada por los datos, durante la cual se comprimen los datos de entrada iniciales en un conjunto de sub-clústeres que se puede administrar. El segundo paso, utiliza un algoritmo de segmentación jerárquico para fundir progresivamente los sub-clústeres en clústeres cada vez más grandes sin necesidad de realizar otra pasada por los datos. La agrupación en clústeres jerárquica tiene la ventaja de que no es necesario seleccionar el número de clústeres por adelantado.

1.4.7.3 Algoritmos metaheurísticos

Aunque entraremos con más profundidad en estos algoritmos en el siguiente apartado, realizamos una primera aproximación a ellos como una alternativa al problema de segmentación abordado en esta investigación.

Otro enfoque utilizado en los problemas de segmentación, es tratar a éste como un problema de optimización que localiza los centroides óptimos de los clústeres en lugar de encontrar la partición óptima. Este enfoque permite la utilización de algoritmos metaheurísticos para problemas de optimización.

En primer lugar definamos qué se entiende por un problema de optimización, por un óptimo global y por un óptimo local (134).

Sea S un espacio de búsqueda, $\Omega \subseteq S$ un subconjunto factible de S y f la función objetivo (también conocida como la función de aptitud o ajuste). Entonces, el **problema de optimización** equivale a encontrar un $\vec{X} \in \Omega$ tal que:

$$f(\vec{X}) \leq f(\vec{Y}) \quad \forall \vec{Y} \in \Omega \quad (\text{para problemas de minimización})$$

$$f(\vec{X}) \geq f(\vec{Y}) \quad \forall \vec{Y} \in \Omega \quad (\text{para problemas de maximización})$$

En este caso, se dice que \vec{X} es el **óptimo global** de f en S .

La función objetivo f puede ser numérica ($f: S \rightarrow \mathcal{R}$) u ordinal ($f: S \times S \rightarrow \mathcal{R}$)

Óptimo local: sea d una distancia definida entre dos soluciones de S tal que $d: S \times S \rightarrow \mathcal{R}$, entonces $\forall \vec{X} \in S$ un vecindario de \vec{X} puede ser definido como:

$$N(\vec{X}) = \{ \vec{Y} \in S \mid d(\vec{X}, \vec{Y}) \leq \varepsilon \} \text{ con } \varepsilon \in \mathcal{R}$$

Se dice que una solución $\vec{X} \in \Omega$ es un **mínimo local** si $f(\vec{X}) \leq f(\vec{Y}) \quad \forall \vec{Y} \in N(\vec{X})$.

Análogamente, se dice que una solución $\vec{X} \in \Omega$ es un **máximo local** si

$$f(\vec{X}) \geq f(\vec{Y}) \quad \forall \vec{Y} \in N(\vec{X}).$$

La bibliografía sobre optimización es muy amplia y en ella se encuentran multitud de métodos y clasificaciones. En este sentido, con la clasificación mostrada en la Figura 71 sobre algoritmos de optimización (134) no se pretende revisar exhaustivamente todos los métodos sino representar esquemáticamente las diferentes clases de métodos de optimización para situar la técnica que usaremos en esta investigación.

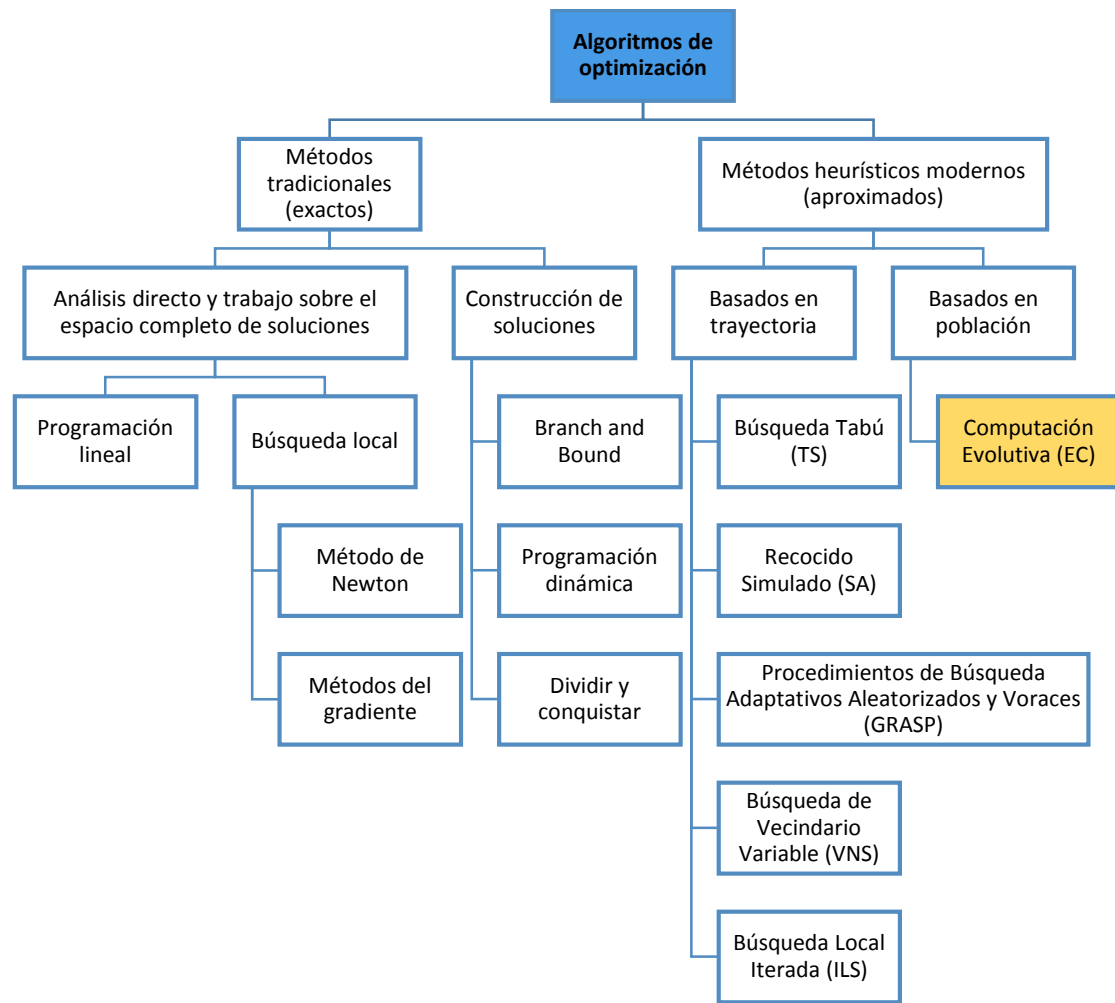


Figura 71. Clasificación de algoritmos de clasificación (adaptado y ampliado de (134))

A continuación, por interés de esta investigación, el grupo de algoritmos de Computación Evolutiva se expande como puede observarse en la Figura 72.

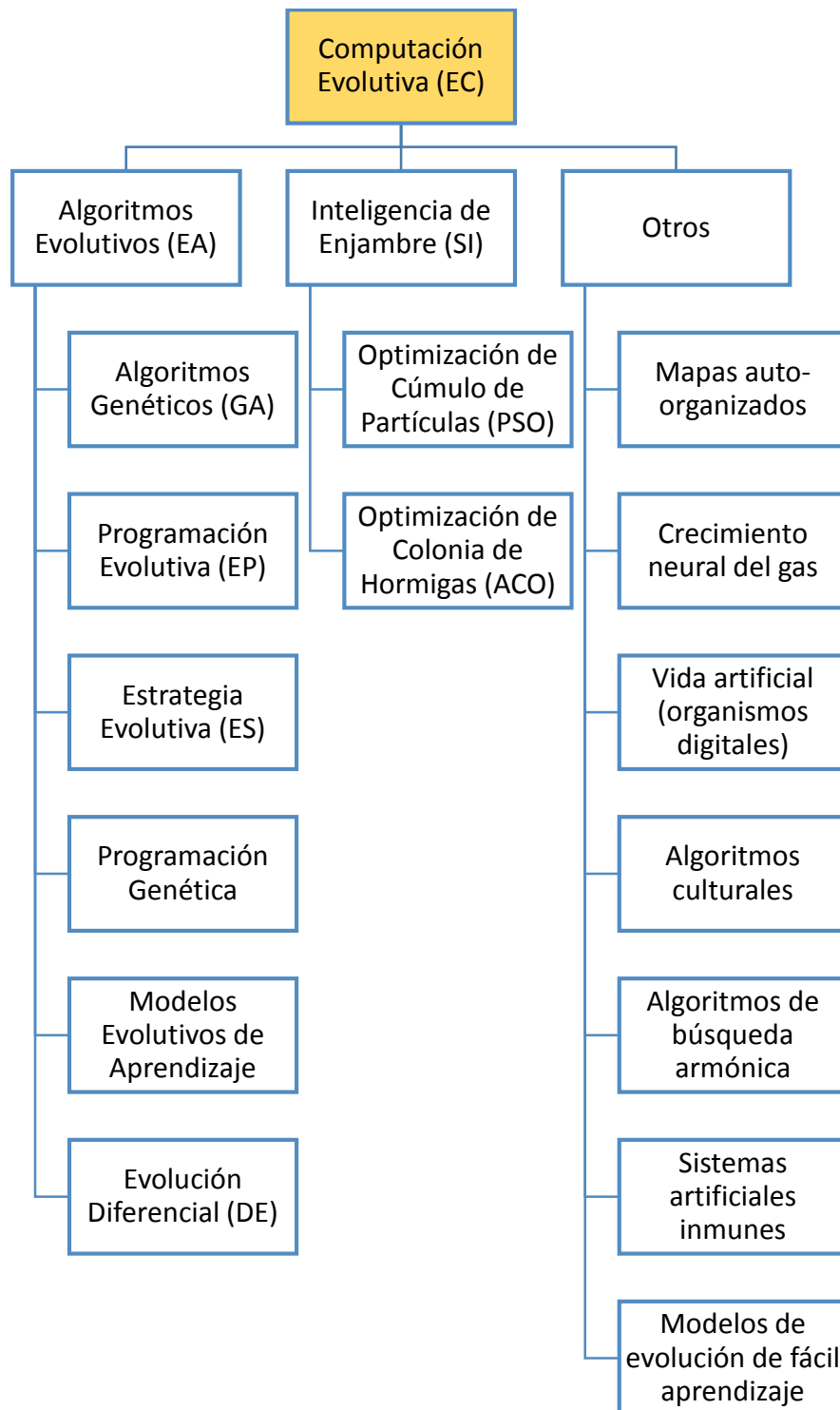


Figura 72. Clasificación de los algoritmos de Computación Evolutiva

En el siguiente apartado se mostrarán las características principales de los algoritmos metaheurísticos.

Un método heurístico es un “procedimiento simple, a menudo basado en el sentido común, que se supone que ofrecerá una buena solución (aunque no necesariamente la óptima) a problemas difíciles, de un modo fácil y rápido” (Zanakis y Evans, 1981).

Las técnicas metaheurísticas son procedimientos de búsqueda de soluciones que, como las heurísticas, no garantizan la obtención de la solución óptima del problema considerado y que también se basan en la aplicación de reglas relativamente sencillas. A diferencia de los heurísticos, las técnicas metaheurísticas tratan de huir de óptimos locales orientando la búsqueda en cada momento dependiendo de la evolución del proceso de búsqueda.

El término “metaheurística” fue introducido por Glover (1971) que define las técnicas metaheurísticas como métodos que integran procedimientos de mejora local y estrategias de alto nivel para crear un proceso capaz de escapar de los óptimos locales y realizar en el espacio de búsqueda total una búsqueda robusta. En su evolución, estos métodos han incorporado diferentes estrategias para evitar la convergencia a óptimos locales, especialmente en espacios de búsqueda complejos.

Los algoritmos metaheurísticos son unos procedimientos iterativos cuya idea básica es la de combinar de forma inteligente diferentes métodos heurísticos a un nivel más alto para conseguir una exploración y una explotación del espacio de búsqueda de soluciones de una forma efectiva y eficiente. Son algoritmos aproximados de optimización y búsqueda de propósito general. Las técnicas metaheurísticas son capaces de proporcionar muy buenas soluciones, no necesariamente las óptimas pero sí aproximadas, en tiempo y recursos razonables.

De la revisión de la literatura sobre técnicas metaheurísticas se pueden encontrar ciertas propiedades fundamentales que las caracterizan:

- El objetivo es una búsqueda eficiente que encuentre soluciones óptimas o “casi óptimas”.

- El procedimiento de cualquier metaheurística es genérico, no depende del problema. La búsqueda no es guiada por la naturaleza específica del problema sino por una estrategia de búsqueda de alto nivel.
- Usualmente, no son algoritmos determinísticos y usan diferentes mecanismos para no quedar atrapados en óptimos locales.
- Incorporan mecanismos para evitar la exploración en regiones del espacio de búsqueda no óptimas.
- Las metaheurísticas utilizan funciones de bondad para cuantificar el grado de adecuación de una determinada solución.

Como resumen podemos decir que una metaheurística es una estrategia genérica de alto nivel que usa diferentes métodos heurísticos de exploración para la búsqueda, en espacios de búsqueda de gran tamaño, de una solución óptima o “casi óptima”. La técnica metaheurística debe identificar de forma rápida las regiones prometedoras del espacio de búsqueda global y no derrochar tiempo en regiones que hayan sido exploradas y no contengan soluciones de calidad.

Las tendencias avanzadas en el uso de las técnicas metaheurísticas (172) usan mecanismos para mejorar el proceso de búsqueda logrando un buen balance entre la diversificación y la intensificación. El término intensificación se refiere a la explotación de algún área prometedora del espacio de búsqueda, mientras que la diversificación se refiere a la exploración para identificar estas áreas prometedoras. Blum y Roli (173) describen la importancia de estos dos conceptos en su estudio del arte acerca de metaheurísticas.

Las diferentes estrategias de exploración del espacio de búsqueda, seguidas por las técnicas metaheurísticas, se diseñan balanceando estática o dinámicamente el aprovechamiento de la experiencia acumulada de búsqueda (intensificación) y la exploración del espacio de búsqueda (diversificación). En este tipo de técnicas es especialmente importante el correcto equilibrio (generalmente dinámico) que haya entre diversificación e intensificación. En este contexto “diversificación” se refiere a la evaluación de soluciones en regiones alejadas del espacio de búsqueda (de acuerdo a

una distancia previamente establecida entre soluciones); también se conoce como exploración del espacio de búsqueda. Por otro lado, “intensificación” hace referencia a la evaluación de soluciones en regiones acotadas y pequeñas con respecto al espacio de búsqueda centradas en el vecindario de soluciones concretas (explotación del espacio de búsqueda). El equilibrio entre estos dos aspectos contrapuestos es de gran importancia, ya que por un lado deben identificarse rápidamente las regiones prometedoras del espacio de búsqueda global y por otro lado no se debe malgastar tiempo en las regiones que ya han sido exploradas o que no contienen soluciones de alta calidad.

Siarry y Michalewicz (1976) identifican características comunes que se presentan claramente en la mayoría de las metaheurísticas; como el uso de la diversificación, forzar la exploración de regiones del espacio de búsqueda, y el uso de la intensificación para investigar a fondo algunas regiones prometedoras. Otra característica común es el uso de la memoria para almacenar las mejores soluciones encontradas. Por último, un defecto común identificado por estos autores en la mayoría de los algoritmos metaheurísticos es el ajuste adecuado de sus parámetros.

Aunque hay varias formas de clasificar los métodos heurísticos, en la Figura 71 las técnicas metaheurísticas probabilísticas se han dividido, por ser la clasificación más usual, en metaheurísticas basadas en trayectoria y basadas en población.

Las metaheurísticas basadas en población se caracterizan por trabajar con un conjunto de soluciones (población) en cada iteración, a diferencia de los métodos anteriormente descritos que únicamente utilizan un punto del espacio de búsqueda por iteración. El resultado final proporcionado por este tipo de algoritmos depende de la forma en que manipula la población. Como veremos más adelante, en esta investigación se ha optado por un algoritmo basado en población.

A continuación, se describen someramente las características principales de los algoritmos metaheurísticos más conocidos para una mejor diferenciación con el algoritmo principal utilizado en esta investigación.

1.4.7.3.1 Algoritmos metaheurísticos basados en trayectoria

En las metaheurísticas basadas en trayectoria su principal característica es que parten de un punto y mediante la exploración de su vecindario, van modificando la solución actual formando una trayectoria en el espacio de soluciones. Por lo tanto, con este método una solución sucesora puede o no pertenecer al vecindario de la solución actual.

La mayoría de estos métodos han surgido como una extensión de los métodos de búsqueda local simple a los que se les ha añadido alguna característica para no quedar atrapados en los óptimos locales. Esto implica la necesidad de una condición de parada más compleja que la de encontrar un óptimo local. La búsqueda finaliza normalmente cuando se alcanza un número máximo de iteraciones establecido de antemano, se encuentre una solución con una calidad aceptable o se detecte un estancamiento del proceso.

Entre estos algoritmos destacan los siguientes:

1. Recocido Simulado (*Simulated Annealing* - SA)

El algoritmo de enfriamiento simulado o *Simulated Annealing* (SA) es considerado como uno de los precursores entre los algoritmos metaheurísticos y uno de los primeros en explicitar una estrategia para escapar de los óptimos locales. Este método se presentó como un algoritmo de búsqueda por Kirkpatrick y col., en 1983 (174) y Cerney, en 1985 (175); cada uno de forma independiente. El algoritmo se origina de un mecanismo estadístico denominado “metrópolis” (176). Los conceptos del SA se inspiran en el proceso de enfriamiento físico de metales y cristales. En cada iteración del algoritmo, se comparan los valores objetivos de la solución actual S y una solución S' del vecindario $N(S)$. Se acepta S' como mejor solución que S si el valor de la función objetivo (o *fitness*) en S' es mejor que el valor de la función objetivo en S ; en cuyo caso S' pasa a ser ahora la solución actual en la nueva iteración. Por otro lado, si S' tiene un valor de la función objetivo inferior que en S , se le asigna una probabilidad de ser aceptada con la esperanza de escapar de un óptimo local. La probabilidad de aceptar una

solución de inferior calidad depende del parámetro de temperatura T y de la variación de la función objetivo $f(S') - f(S)$ entre ambas soluciones (para el caso de minimización). Habitualmente, T se decrementa con cada iteración del algoritmo.

2. Búsqueda Tabú (*Tabu Search* - TS)

La búsqueda tabú o *Tabu Search* (TS) es una de las metaheurísticas más citadas y empleadas con más éxito a la hora de resolver problemas de optimización combinatoria. La idea básica fue introducida por Glover en 1993 (171). Una buena descripción del método se puede encontrar en (178).

En cada iteración, el algoritmo TS explora el espacio de búsqueda al moverse desde la solución actual S a una vecina, donde el vecindario de la solución actual $N(s)$ está formado por transformaciones permitidas de la solución S realizadas por una heurística ad hoc específica (también denominada operador de vecindario). Cuando se encuentra un óptimo local, la búsqueda se mueve a la mejor solución del vecindario, la cual no es necesariamente mejor que S .

Para evitar buscar varias veces en la misma región, se mantiene un historial de búsquedas en una memoria de corto plazo denominada “lista tabú”. En esta lista se almacenan las soluciones visitadas recientemente para excluirlas de los próximos movimientos; a menos que reúnan ciertas condiciones (criterio de aspiración). El criterio de aspiración más ampliamente usado es permitir soluciones cuyo valor en la función objetivo (valor *fitness*) sea mejor que el de la mejor solución encontrada hasta el momento. La lista tabú, por lo general, almacena solo una cantidad limitada y fija de información. Una opción rara vez usada es registrar soluciones completas, porque requiere una cantidad importante de almacenamiento y el alto coste de comprobar si un movimiento potencial es tabú o no. La opción más usada consiste en registrar las últimas transformaciones realizadas de la solución actual y prohibir transformaciones en

reversa; otras están basadas en características claves de las soluciones o de los movimientos.

3. Procedimientos de Búsqueda Adaptativos Aleatorizados y Voraces (*Greedy Randomized Adaptive Search Procedures – GRASP*)

GRASP (179) es un procedimiento iterativo que presenta un diseño relativamente simple, que puede verse en (180). Combina una heurística *ad hoc* constructiva aleatoria-voraz con una búsqueda local. El procedimiento *GRASP* sigue un método multi-arranque diseñado para resolver problemas difíciles en optimización combinatoria. En su versión básica, cada iteración consiste en dos fases: una fase constructiva cuyo producto es una solución factible y buena, aunque no es necesariamente un óptimo local, y otra fase de búsqueda local, durante la cual se examinan vecindades de la solución. Al llegar a un óptimo local la iteración termina.

4. Búsqueda Local Iterada (*Iterated Local Search - ILS*)

La búsqueda local iterada o *Iterated Local Search* (ILS) (181) es una metaheurística basada en un concepto simple pero muy efectivo. Esta técnica se fundamenta en que en cada iteración a la solución actual se le aplica un cambio o modificación que da lugar a una solución intermedia o perturbada. A esta nueva solución se le aplica una heurística base para mejorarla, que suele ser un método de búsqueda local. Este nuevo óptimo local obtenido por el método de mejora puede ser aceptado como nueva solución actual si pasa un test de aceptación.

El proceso de perturbación de la solución tiene una importancia obvia ya que si la perturbación es demasiado pequeña puede que el algoritmo no sea capaz de escapar del óptimo local. Pero si es demasiado grande, la perturbación puede hacer que el algoritmo sea como un método de búsqueda local con un reinicio aleatorio (o re-arranque aleatorio). Por este motivo, el método de perturbación

debe generar una nueva solución que sirva como inicio a la búsqueda local, pero sin que esté muy alejada de la actual para que no sea una solución aleatoria.

5. Búsqueda con Vecindario Variable (*Variable Neighborhood Search* - VNS)

La Búsqueda con Vecindario Variable o *Variable Neighborhood Search* (VNS) es una metaheurística propuesta por Hansen y Mladenović (182), basada en el principio de cambiar de manera sistemática la exploración del vecindario para escapar así de mínimos locales.

El algoritmo VNS define una sucesión de vecindarios $N_1(S)$, $N_2(S)$, ..., $N_k(S)$ y cada iteración del algoritmo contempla tres fases: la elección del candidato, una fase de mejora y la fase del movimiento. En la primera fase, se elige una solución S' del vecindario actual $N_k(S)$ y se mejora, en la segunda fase, mediante una rutina de búsqueda local; esto proporciona un nuevo mínimo local S'' . Si dicho mínimo es menor que el mejor encontrado hasta el momento, es decir, $f(S'') \leq f(s)$, entonces S'' se convierte en la solución actual, inicializándose el contador de vecindarios a 1, y la búsqueda continua desde $N_1(S'')$; en otro caso, la búsqueda continua desde $N_{k+1}(S)$ en la siguiente iteración; es decir, se repite el proceso pero utilizando el siguiente vecindario.

La búsqueda local es el paso de intensificación del método y el cambio de vecindario puede considerarse como el paso de diversificación.

1.4.7.3.2 Metaheurísticas basadas en población

Al comienzo de los años 60, una observación aguda sobre la relación subyacente entre la optimización y la evolución biológica, condujo al desarrollo de un importante paradigma en la inteligencia computacional denominada "Computación Evolutiva" (*Evolutionary Computing*-EC) cuyo objeto es la realización de búsquedas y optimizaciones complejas.

Como parte del proceso de evolución este paradigma, en 1995 apareció un nuevo algoritmo de computación evolutiva conocido como de Evolución Diferencial (*Evolution Differential*-DE), que llamó la atención de los investigadores de todo el mundo por su

alto grado de robustez, velocidad de convergencia y precisión en problemas de optimización de parámetros reales. Su algoritmo, de código sencillo y pocos parámetros, ha superado a algoritmos veteranos de la familia de la Computación Evolutiva (algoritmos genéticos (AG), estrategias evolutivas (EE), y algoritmos meméticos (AM)) en problemas de referencia y del mundo real.

La computación bioinspirada, como también se conoce la computación evolutiva, se basa en emplear analogías con sistemas naturales o sociales para diseñar métodos heurísticos no determinísticos de “búsqueda”, de “aprendizaje”, de imitación de “comportamiento”, etc. En la actualidad, el campo de investigación en el diseño de algoritmos bioinspirados, es uno de los más prometedores por los buenos resultados que están consiguiendo.

Los algoritmos bioinspirados permiten modelar, de forma aproximada, un fenómeno existente en la naturaleza que es utilizado como base para la construcción de una metáfora biológica que se utiliza en la resolución de problemas. Son algoritmos no determinísticos y adaptativos; en el sentido que utilizan realimentación con el entorno para modificar el modelo y sus parámetros.

Veamos ahora las características de los principales algoritmos.

1.4.7.3.2.1 Algoritmos evolutivos

Los algoritmos evolutivos forman parte de la llamada inteligencia artificial (IA) y nacen por la necesidad de obtener nuevos métodos capaces de resolver problemas complejos como alternativa a la optimización heurística. El concepto en el que se basan estos algoritmos es un símil de la evolución natural expuesta por Charles Darwin en su teoría de la evolución (183). La adaptación a contextos de problemas de optimización consiste en que cada individuo se asemeja a una posible solución del problema, que evoluciona interaccionando con otros individuos hasta obtener un resultado suficientemente bueno, es decir, un individuo cuyo material genético (codificación) resuelve de manera aceptable el problema planteado.

Los algoritmos evolutivos fueron introducidos por investigadores desde diferentes ciudades de forma simultánea con un intervalo de tiempo de quince años. Estos fueron: Fogel (183) desde USA, Holland (184) desde Dortmund (al que se le considera uno de los padres de la computación evolutiva), Schwefel (186), que desarrolló la teoría para la estrategia evolutiva y Rechenberg, desde Berlín (187).

Algoritmos genéticos (*Genetic Algorithms* - GA)

John Holland (184) propuso en 1971 una clase atractiva de modelos computacionales que llamó algoritmos genéticos (*Genetic Algorithms*-GA) que imitan el proceso de la evolución biológica para resolver problemas en un amplio dominio.

Dentro de los algoritmos evolutivos, los GA son los más populares. Están inspirados en la evolución natural de las especies de la siguiente forma: en un ecosistema real, con una población de seres vivos (individuos), éstos transmiten su código genético a la siguiente generación mediante la reproducción entre ellos. De esta manera, sólo los más “fuertes” sobrevivirán y serán capaces de reproducirse; transmitiendo a su vez el material genético a sus descendientes. Además, éstos podrán experimentar mutaciones provocadas por el entorno en el que viven, haciendo que el individuo sea más apto para su adaptación al medio y así poder reproducirse y transmitir a la siguiente generación su información genética.

Su objetivo, en base a la metáfora de la evolución de las especies, es obtener individuos lo suficientemente buenos para que vayan pasando de generación en generación

El diagrama de flujo general de un algoritmo genético se observa en la Figura 73.

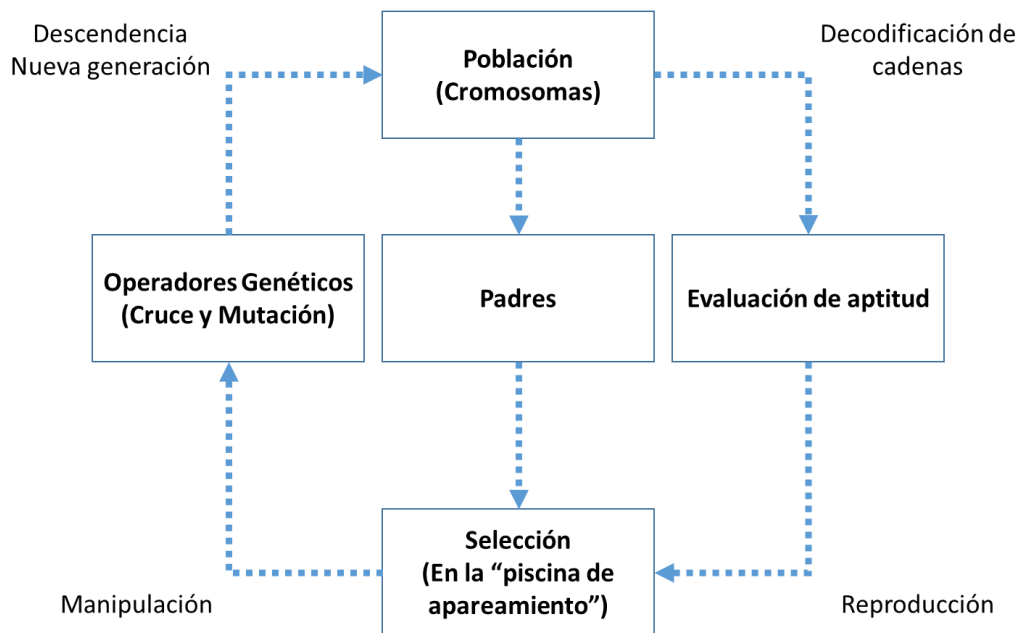


Figura 73. Diagrama de flujo de un GA

La actividad de selección juega un papel fundamental en los GA a la hora de determinar cómo los individuos compiten entre sí por la supervivencia de genes. La selección elimina a las malas soluciones y mantiene a las buenas. La selección de individuos, es un factor determinante en los resultados. Su objetivo es determinar los padres que se encargarán de generar la nueva población. Son los individuos con “capacidad de reproducirse” dentro de nuestra población. Existen varios métodos para llevarla a cabo: método de ruleta (187) (probabilidad de selección en proporción a la aptitud del individuo dado), método de la ruleta extendido, método de torneo (donde un número de individuos aleatorios aleatoriamente escogidos se comparan entre sí) (188), etc.

Debido a la aleatoriedad en la selección de la mayoría de los métodos de selección (134) no se puede garantizar la supervivencia de la mejor solución actual. Tampoco pueden los operadores tradicionales de recombinación o mutación más tradicionales. Sin embargo, el “elitismo” ofrece esta garantía al seleccionar explícitamente el mejor individuo o grupo de individuos. Tanto la selección aleatoria como la selección elitista pueden conducir a duplicados; no obstante, serán buenos candidatos.

La recombinación de los individuos se realiza para investigar el rendimiento de nuevos individuos que se asemejen a los existentes. Esto se realiza a nivel de genotipo de los individuos y conduce a la construcción de nuevas soluciones intermedias. La noción de generaciones surge de los padres que recombinan sus genes para crear descendencia. Por lo general, los padres se apartan para dejar espacio a la descendencia que llevarán algunos de sus genes. La recombinación se hace a menudo a través de una operación de cruce (*crossover*), entre dos descendientes de dos padres dando a cada descendiente una parte de cada gen de su padre.

Por otro lado, la mutación se presenta con menor frecuencia tanto en la naturaleza como en los algoritmos genéticos. La operación de mutación puede mantener la diversidad de la población y contribuir a la exploración del espacio de búsqueda. La exploración por mutación es, a menudo, lenta comparada con la recombinación, pero en problemas donde la recombinación es perturbadora, la mutación puede ser una forma importante para explorar el entorno.

Aunque la mayor parte de la búsqueda de soluciones en los algoritmos evolutivos se realiza a través de la recombinación, la mutación puede ser vital para proporcionar la diversidad que la recombinación necesita (134).

El esquema general de un algoritmo genético es el siguiente:

Begin

Creación de la población inicial P_0 .

Evaluación de P_0 .

While condición de parada no alcanzada (P_i cumpla el criterio de convergencia)

While Seleccionar individuos de la población actual P_i do

Cruzar con cierta probabilidad obteniendo descendientes.

Mutar los descendientes con cierta probabilidad.

Evaluar los nuevos individuos generados.

Añadir los nuevos individuos a la población auxiliar P_i^0 .

Endw

Combinar los individuos de P_i y P_i^0 según cierta estrategia para generar P_{i+1} .

Evaluar población P_{i+1} .

Endw

End

Para más detalles del Algoritmo Evolutivo puede consultarse, entre otros, la publicación de Michalewicz y Fogel (1992).

Programación evolutiva (*Evolutionary Programming* - EP)

La programación evolutiva (ES) fue propuesta por Fogel en los años 60 (190). En este paradigma de optimización, la inteligencia es vista como un comportamiento adaptativo. Fogel utilizó la "programación evolutiva" [15] para hacer evolucionar autómatas de estados finitos, de manera que fuesen capaces de predecir secuencias futuras de símbolos que recibirían. Fogel utilizó una función de recompensa para indicar si un autómata era bueno o no para predecir un cierto símbolo.

El esquema general de un algoritmo de programación evolutiva es el siguiente:

Begin

Generar aleatoriamente una población

Evaluar la aptitud de la población

Repetir

Aplicar operador de mutación a cada individuo de la población

Evaluar cada hijo resultante de la mutación

Realizar la selección mediante torneo entre el padre y el hijo

Hasta que se cumpla la condición de paro

End

En la programación evolutiva no se aplica ningún operador de recombinación, pues la simulación es a nivel de las especies (las especies distintas no se pueden recombinar para crear nuevos individuos).

Estrategia Evolutiva (*Evolution Strategy* - ES)

En 1963, dos estudiantes de la Universidad Técnica de Berlín, Rechenberg (191) y Schwefel (192) desarrollaron las estrategias evolutivas como un método de ajustes discretos aleatorios inspirado en el mecanismo de mutación que ocurre en la naturaleza.

La versión original $(1+1)$ -EE usaba un solo padre y con él se generaba un solo hijo. Este hijo se mantenía si era mejor que el padre, o de lo contrario se eliminaba (a este tipo de selección se le llama extintiva, porque los peores individuos tienen una probabilidad igual a cero de ser seleccionados). Rechenberg (191) introdujo el concepto de población, al proponer una estrategia evolutiva llamada $(\mu + 1)$ - EE, en la cual hay μ padres y se genera un solo hijo, el cual puede reemplazar al peor padre de la población (selección extintiva). Schwefel (192) introdujo en 1975 el uso de múltiples hijos en las denominadas $(\mu + \lambda)$ - EEs y (μ, λ) - EEs. La notación se refiere al mecanismo de selección utilizado:

- En el primer caso, los μ mejores individuos obtenidos de la unión de padres e hijos sobreviven.
- En el segundo caso, sólo los μ mejores hijos de la siguiente generación sobreviven.

Rechenberg (191) formuló una regla para ajustar la desviación estándar de forma determinística durante el proceso evolutivo de tal manera que el procedimiento convergiera hacia el óptimo. Esta regla se conoce con el nombre de “regla del éxito 1/5”, y hace referencia al razonamiento siguiente: “la razón entre mutaciones exitosas y el

total de mutaciones debe ser $1/5$. Si es más, entonces debe incrementarse la desviación estándar. Si es menos, entonces debe decrementarse”.

El algoritmo básico de la estrategia evolutiva es el siguiente:

- Inicialización: generación aleatoria de una población inicial.
- Variación: operadores de mutación y recombinación.
- Evaluación: aptitud (*fitness*) de cada individuo.
- Selección: determinística frente a la selección estocástica de la programación evolutiva y algoritmos genéticos.

Las diferencias entre los algoritmos genéticos y la estrategia evolutiva, representadas por Hernández y col. (192), pueden verse en la Figura 74.

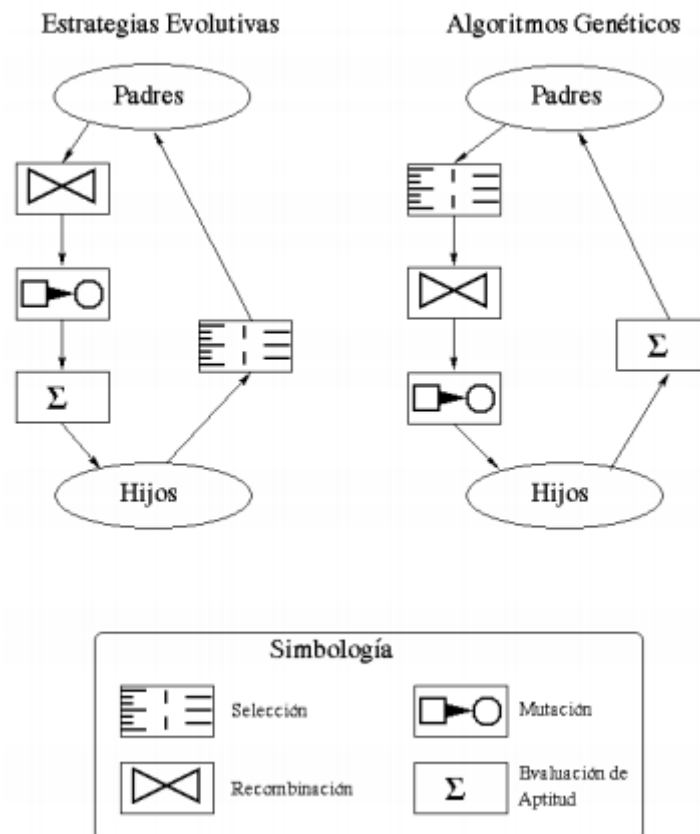


Figura 74. Representación conceptual de las EEs y los AGs.

Programación Genética (*Genetic Programming* - GP)

John R. Koza (191) es considerado como el "padre" de la programación genética quien acuñó el término "programación genética" en 1992. Aun así, sus antecedentes se remontan a 1954, cuando Nils Aall (192) aplicó algoritmos evolutivos a simulaciones evolutivas, y a 1985, año en el que Michael L. Cramer estableció los primeros avances en el uso de los árboles de decisión en la programación genética. Fue definitivamente John R. Koza quien lo aplicó a la resolución de varios problemas complejos de optimización y búsqueda.

La programación genética puede ser considerada como un paradigma de computación evolutiva para encontrar programas de ordenador que realicen una tarea definida por el usuario. Se trata de una especialización de los algoritmos genéticos donde cada individuo es un programa de ordenador. Por tanto, puede considerarse una técnica de aprendizaje automático usada para optimizar una población de programas de ordenador según una heurística definida en función de la capacidad del programa para realizar una determinada tarea computacional definida por el usuario.

La programación genética requiere grandes exigencias desde el punto de vista computacional, razón por la cual en principio solo se aplicaba a problemas relativamente sencillos. Hoy día, con la mejora de elementos hardware y del propio algoritmo es posible resolver problemas de mayor complejidad.

La principal diferencia entre la programación genética y los algoritmos genéticos es la representación de la solución. La programación genética crea programas informáticos, en LISP o lenguajes de programación, como la solución. Los algoritmos genéticos crean una serie de números que representan la solución.

Los cinco pasos previos (195) para la versión básica de la programación genética requiere que el usuario especifique la siguiente información:

1. El conjunto de terminales para cada rama del programa a ser evolucionado.
2. El conjunto de funciones primitivas para cada rama del programa a evolucionar.
3. La medida de la aptitud o *fitness*.

4. Selección de los parámetros para el control de la ejecución.
5. El criterio de terminación y el criterio para designar el resultado.

Según Koza (193), la programación genética utiliza cuatro pasos para resolver problemas:

1. Generar una población inicial de composiciones aleatorias de las funciones y terminales del problema (programas informáticos).
2. Ejecutar cada programa en la población y asignarle un valor físico de acuerdo con lo bien que resuelve el problema.
3. Crear una nueva población de programas de ordenador.
 - Copiar los mejores programas existentes.
 - Crear nuevos programas informáticos por mutación.
 - Crear nuevos programas informáticos de cruce (reproducción sexual).
4. El mejor programa de ordenador que apareció en cualquier generación, la mejor solución, se designa como el resultado de la programación genética.

En la Figura 75 puede verse una representación de las operaciones de mutación y cruce en este tipo de algoritmos.

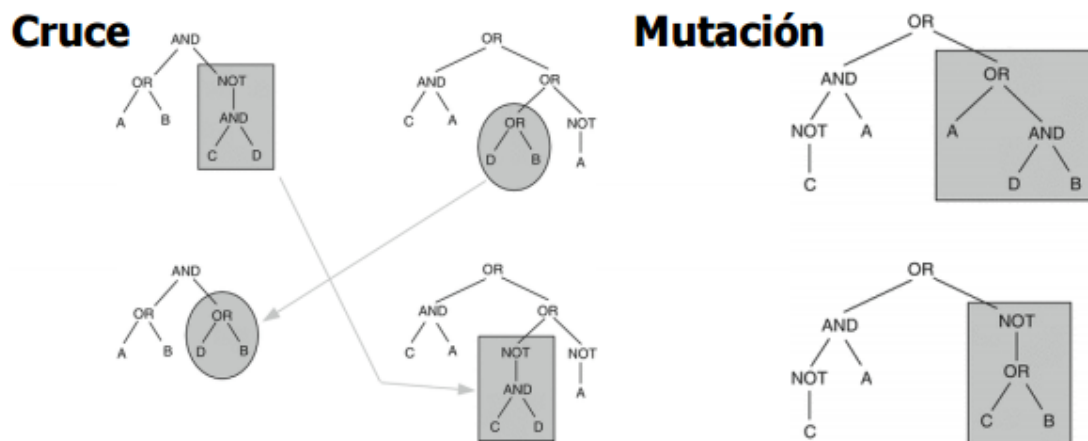


Figura 75. Operaciones de cruce y mutación en programación genética

Modelos evolutivos de aprendizaje

El objetivo de un proceso de aprendizaje es obtener una cierta regla o sistema que permita clasificar objetos de un modo automático, o predecir el valor de las variables de control de un sistema. El problema fundamental de la predicción está en modelar la relación entre las variables de estado para obtener el valor de la variable de control. Mientras que el problema fundamental de la clasificación está directamente relacionado con la separabilidad de las clases.

Podemos encontrar diferentes formas de uso de los algoritmos evolutivos en problemas de aprendizaje y extracción de conocimiento en las bases de datos:

- Aprendizaje evolutivo de sistemas basados en reglas (reglas “intervalares” para tratar con datos continuos, reglas *fuzzy*,...). Los cromosomas codifican una regla o un conjunto de ellas y cada gen codifica una regla o parte de una regla. Existen cuatro modelos generales que están fundamentados en el uso de los algoritmos evolutivos para aprender sistemas basados en reglas (*Genetics Based Machine Learning*).
 - Cromosoma = Base de Reglas.
 - Modelo Pittsburgh: GASSIST, CORCORAN, GIL.
 - Cromosoma = Regla.
 - Modelo Michigan (XCS, UCS) LCS - *Learning Classifier Systems*.
 - Modelo de Aprendizaje Iterativo IRL – *Iterative Rule Learning* (SIA, HIDER).
 - Modelo de Cooperación-Competición GCCL – *Genetic Cooperative-Competitive Learning* (REGAL, LOGEMPRO).
- Aprendizaje con programación genética para regresión y clasificación.
- Aprendizaje evolutivo híbrido (redes neuronales evolutivas, extracción evolutiva de prototipos, *clustering evolutivo*...).

- Aplicación en diferentes áreas de extracción de conocimiento en bases de datos (*Knowledge Discovery in Databases* - KDD): reducción de datos, extracción de reglas de asociación en problemas de *data mining*, etc.

Evolución diferencial (*Differential Evolution* - DE)

El concepto de evolución diferencial (DE) ha surgido recientemente como algoritmo simple y eficiente para la optimización global en espacios continuos. La evolución diferencial comparte muchas características con los algoritmos genéticos (GAs), aunque es mucho más fácil de implementar que los GAs y aplica un tipo de operador de mutación diferencial sobre los cromosomas de los padres para generar la descendencia. Desde su creación en 1995, la evolución diferencial ha llamado la atención de muchos investigadores en todo el mundo, dando lugar a una gran cantidad de variantes del algoritmo básico con un rendimiento mejorado. Este será el algoritmo utilizado en esta investigación y además de en este apartado, será tratado específicamente en el siguiente.

El algoritmo de evolución diferencial (DE), es original de los investigadores Price y Storn en 1995 (190). En un primer momento el DE fue desarrollado entorno a la idea de encontrar métodos que resolvieran el problema de ajuste del polinomio de *Tchebychev*. Para ello, se basaron en una variación del método llamado *genetic annealing* (recocido genético). Sus primeros logros en el mundo de los algoritmos evolutivos fue en el “*First International Contest on Evolutionary Optimization*”, donde obtuvo el tercer puesto como mejor método de optimización. Desde entonces, es un algoritmo que siempre consigue magníficos resultados computacionales en los congresos especializados.

Con objeto de hacer más clara la exposición, veamos la notación y los términos que utilizaremos en ella y de acuerdo con la seguida por Das (134).

DE realiza la búsqueda de un óptimo global en un espacio D-dimensional. Para ello, comienza con una población inicial aleatoria de NP vectores de parámetros reales D-dimensionales.

Cada vector, también conocido como “genoma” o “cromosoma”, forma una solución candidata para el problema de optimización multidimensional.

Las diferentes iteraciones del problema se conocen como “generaciones” y las representaremos como $t=0, 1, 2, \dots, t+1, \dots$ (en la literatura existente también se utiliza $G=0, 1, 2, \dots$ y $g=0, 1, 2, \dots$). Dado que los vectores de parámetros son susceptibles de ser cambiados a lo largo de diferentes generaciones, adoptamos la siguiente notación para representar el vector i -ésimo de la población en la generación actual (es decir en $t = t$) como:

$$\vec{X}_i(t) = [x_{i,1}(t), x_{i,2}(t), \dots, x_{i,D}(t)]^T$$

donde $i=1, 2, \dots, NP$. $\vec{X}_i(t)$ se conoce como “vector objetivo” (*target vector*).

Aunque el término “mutación”, desde el punto de vista biológico, expresa un cambio repentino en las características genéticas de un cromosoma, en el contexto de la Computación Evolutiva la mutación es entendida como un cambio o perturbación en un vector en cuya formación se aplica un efecto aleatorio.

En el caso de los algoritmos DE, la mutación de dos vectores produce un nuevo vector que se denomina “vector donante” (*donor vector*) y se denota por $\vec{V}_i(t)$. Como veremos más adelante, para crear $\vec{V}_i(t)$ necesitaremos tres vectores y se especificará como mutación diferencial (por la diferencia entre dos vectores).

Con el objeto de aumentar la diversidad de la población, una vez que se ha generado el vector donante, entra en escena una nueva operación que se aplica a dicho vector y que se denomina “cruce” (*crossover*). Esta operación de cruce crea un nuevo vector que se denomina “vector de ensayo” (*trial vector*) que se denotará por $\vec{U}_i(t)$.

A modo de resumen podemos decir que:

- Un vector padre de la generación corriente o actual se denomina “**vector objetivo**”.
- Un vector mutante, obtenido mediante una operación de mutación diferencial, se denomina “**vector donante**”.

- Y por último, el vector hijo o descendiente, formado por el cruce o recombinación del vector donante con el vector objetivo, se denomina “**vector de ensayo**”.

El algoritmo DE es una técnica no determinista basada en la evolución de una población de vectores (individuos) de valores reales que representan las soluciones en el espacio de búsqueda. Utilizaremos indistintamente los términos “individuos” y “vectores” según el contexto para una mejor comprensión de la idea que se pretenda expresar.

Los algoritmos de evolución diferencial (DE) operan siguiendo los mismos pasos computacionales que los algoritmos evolutivos (EA). Sin embargo, a diferencia de los EA tradicionales, DE emplea la diferencia de vectores de parámetros (individuos) para explorar la función objetivo (*fitness*) en el espacio de búsqueda.

Al igual que otras técnicas de búsqueda basadas en población, DE genera nuevos vectores (*trial solutions*) como resultados de perturbaciones sobre vectores existentes y lo hace utilizando la diferencia proporcional entre dos vectores de la población actual seleccionados al azar. Para producir el nuevo individuo o descendiente, en su forma más simple DE añade la diferencia proporcional anterior a un tercer vector de la población seleccionado también al azar.

En la etapa de selección, el vector de ensayo compite contra el vector objetivo del mismo índice. Una vez que el último vector de ensayo ha sido testado, los supervivientes de todas las competiciones por parejas se convierten en permanentes para la próxima generación en el ciclo evolutivo.

DE es un algoritmo evolutivo simple que funciona a través de un ciclo simple de cuatro etapas como el que se observa en la Figura 76.

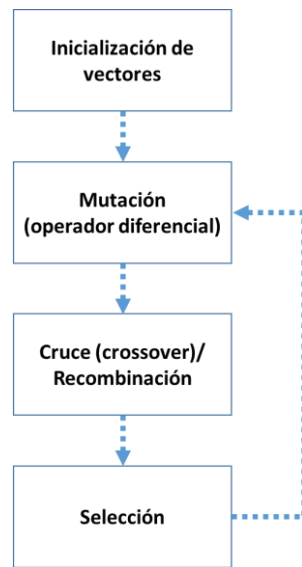


Figura 76. Etapas del algoritmo DE

Veamos a continuación las operaciones principales en los algoritmos de evolución diferencial.

Inicialización

Dado que cada parámetro del problema tendrá un rango de variación concreto, al inicio de la ejecución del algoritmo ($t=0$) se inicializan dentro de su rango.

Por lo tanto, si el parámetro j -ésimo del problema tiene su límite inferior en $x_{min,j}$ y su límite superior en $x_{max,j}$ y sea $rand_{i,j}(0,1)$ la instanciación j -ésima de una distribución uniforme (0,1) para el vector i -ésimo, entonces se puede inicializar el componente j -ésimo del individuo i -ésimo de la población de la siguiente forma (134):

$$x_{i,j}(0) = x_{min,j} + rand_{i,j}(0,1) \cdot (x_{max,j} - x_{min,j})$$

Mutación

A diferencia de lo que ocurre en los algoritmos genéticos (GA) el objetivo de la mutación en los algoritmos de evolución diferencial (DE) es generar nueva descendencia. En concreto, consiste en generar tantos individuos nuevos como el número de individuos que existen la población actual (P_0). Para generar cada uno de estos individuos se seleccionan cuatro vectores: $\vec{X}_i(t)$, $\vec{X}_{r_1^i}(t)$, $\vec{X}_{r_2^i}(t)$, y $\vec{X}_{r_3^i}(t)$.

El individuo $\vec{X}_i(t)$, conocido como “vector objetivo”, es el individuo i -ésimo de la población correspondiente a la generación t , que se quiere sustituir por el nuevo descendiente. Los otros tres individuos son seleccionados aleatoriamente y servirán para generar el nuevo descendiente a través de la expresión siguiente:

$$\vec{V}_i(t) = \vec{X}_{r_1^i}(t) + F \cdot (\vec{X}_{r_2^i}(t) - \vec{X}_{r_3^i}(t))$$

donde F es un factor de mutación que toma valores entre 0 y 1 y que establece el rango de diferenciación entre los individuos r_2^i y r_3^i con el objetivo de evitar el estancamiento en el proceso de búsqueda. Los índices r_1^i , r_2^i y r_3^i son números enteros diferentes entre sí y son elegidos aleatoriamente del intervalo $[1, NP]$. Por último, $\vec{V}_i(t)$ es el conocido “vector donante”.

Cruce

Tras obtener el vector donante, se realiza un cruce entre el vector objetivo y el vector donante. Para ello, el vector donante intercambia parte de sus componentes (o genes) con el vector objetivo, formándose el vector de ensayo $\vec{U}_i(t) = [u_{i,1}(t), u_{i,2}(t), \dots, u_{i,D}(t)]^T$.

En los algoritmos DE existen dos tipos de esquemas u operadores para realizar el intercambio de genes:

- Operador binomial.

El operador binomial selecciona, en función de la probabilidad de cruce (CR), el gen de uno u otro vector como se muestra en la siguiente ecuación:

$$u_{i,j}(t) = \begin{cases} v_{i,j}(t), & \text{si } (rand_{i,j}(0,1) \leq CR \text{ o } j = j_{rand}) \\ x_{i,j}(t), & \text{en otro caso} \end{cases}$$

donde:

- i identifica al individuo i -ésimo de la población.
- j es la posición del gen o componente.
- D es el tamaño del vector del genoma.

- CR es la probabilidad con la que los elementos del vector donante se intercambian por los del vector objetivo. También llamada “probabilidad de cruce”, aparece como un parámetro de control en el algoritmo DE tal y como el parámetro F (factor de mutación).
- Y finalmente, $j_{rand} \in [1, 2, \dots, D]$ es un índice elegido aleatoriamente entre 1 y D, el cual asegura que $\vec{U}_i(t)$ recibe al menos un gen de $\vec{V}_i(t)$.

Un ejemplo del funcionamiento de este operador puede verse en la Tabla 47.

| Componentes de $\vec{X}_i(t)$ | Componentes de $\vec{V}_i(t)$ | $rand_{i,j} (0,1)$ | j_{rand} | CR | Componentes de $\vec{U}_i(t)$ |
|----------------------------------|----------------------------------|--------------------|------------|------|----------------------------------|
| 8,07 | 3,17 | 0,66 | 2 | 0,75 | 3,17 |
| 7,29 | 0,74 | 0,75 | | | 0,74 |
| 8,55 | 1,08 | 0,78 | | | 8,55 |
| 6,47 | 9,02 | 0,64 | | | 9,02 |

Tabla 47. Ejemplo de cruce con operador binomial

- Operador exponencial.

El operador exponencial trabaja de forma diferente al binomial. En primer lugar, se selecciona un número n, entero y aleatorio en el rango $[0, D-1]$, donde D representa el tamaño del vector. Este número determina el punto de inicio en el vector objetivo a partir del cual se seleccionan los elementos del vector donante para formar el vector de ensayo $\vec{U}_i(t)$. A continuación, se selecciona aleatoriamente otro número entero L, perteneciente al intervalo $[1, D]$ y en función de la probabilidad de cruce CR, de acuerdo al siguiente esquema:

L=0

Do mientras $((rand_{i,j} (0,1) < CR) \text{ y } (L < D))$

{

 L=L+1;

};

Se puede comprobar que $P(L \geq a) = CR^{a-1}$ para toda $a > 0$.

El valor de L determinará el número de elementos seleccionados del vector donante que sustituirán a los del vector objetivo. Para cada vector donante deben ser elegidos al azar, con el procedimiento descrito, un nuevo par de números n y L .

El operador exponencial selecciona, en función de n y L , el gen de uno u otro vector como se muestra en la siguiente ecuación:

$$u_{i,j}(t) = \begin{cases} v_{i,j}(t), & \text{para } j = \langle n \rangle_D, \langle n + 1 \rangle_D, \dots, \langle n + L - 1 \rangle_D \\ x_{i,j}(t), & \text{para el resto de } j \in [0, D - 1] \end{cases}$$

donde $\langle n \rangle_D$ es la función módulo de módulo D .

Selección

Por último, en esta etapa se decide cuál de los dos vectores, el objetivo $\vec{X}_i(t)$ o el de ensayo $\vec{U}_i(t)$, sobrevivirá a la siguiente generación; basándose en el concepto de la supervivencia del más apto.

Si el vector de ensayo obtiene un mejor valor en la función objetivo, reemplazará al vector objetivo y será el vector de ensayo quien pase a la siguiente generación. En caso contrario, se descartará al vector de ensayo y será el vector objetivo quien pase a la siguiente generación.

El mejor valor en la función objetivo será el mínimo en el caso de un problema de minimización y el máximo en el de un problema de maximización.

La siguiente ecuación muestra la operación matemáticamente:

$$\vec{X}_i(t+1) = \begin{cases} \vec{U}_i(t), & \text{si } f(\vec{U}_i(t)) \leq f(\vec{X}_i(t)) \\ \vec{X}_i(t), & \text{en caso contrario} \end{cases}$$

Siendo f la función de objetivo o función de aptitud (*fitness*) del problema de optimización.

El pseudocódigo del algoritmo de Evolución Diferencial quedaría de la siguiente forma:

Begin

Creación de la población inicial P_0 .

Evaluación de P_0 .

Mientras no termine do (P_i cumpla el criterio de convergencia)

Mientras Seleccionar individuos de la población actual P_i do Mutar en función de un factor obteniendo descendientes.

Cruzar los descendientes con cierta probabilidad.

Evaluar los nuevos individuos generados.

Si El descendiente tiene mejor *fitness* que su progenitor *entonces*

Reemplazar al progenitor por el descendiente en P_{i+1} .

Si no

Introducir al progenitor en P_{i+1} .

Fin Si

Fin mientras

Fin mientras

End

Este algoritmo aparece gráficamente, con la notación de Price y Storn (190), en la Figura 77.

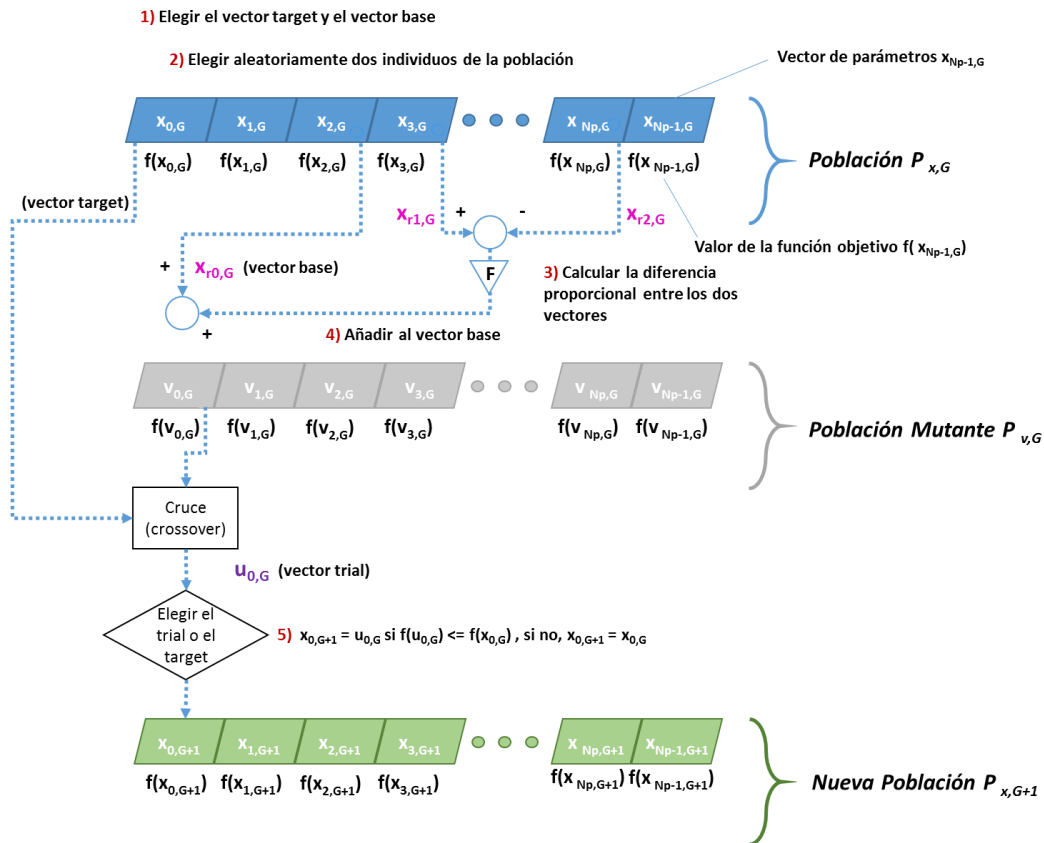


Figura 77. Algoritmo DE

1.4.7.3.2.2 Algoritmos de inteligencia de enjambre

Optimización de Cúmulo de Partículas (*Particle Swarm Optimization - PSO*)

Los algoritmos basados en nubes de partículas (197), (198) y (199), en inglés *Particle Swarm Optimization* (PSO), son también algoritmos metaheurísticos. Estos algoritmos son muy recientes y novedosos, y pertenecen al grupo de los algoritmos bioinspirados (200).

En 1995, Kennedy y Eberhart (197), considerados los padres de estos algoritmos, desarrollaron el algoritmo original experimentando con algoritmos que modelaban el comportamiento del vuelo de algunos pájaros o el movimiento de los bancos de peces.

Cada individuo en el algoritmo PSO, que se denomina partícula, se va moviendo en un espacio multidimensional que representa su espacio social. Debido a su planteamiento, este tipo de algoritmo se adapta muy bien a problemas matemáticos tanto de carácter continuo como de naturaleza discreta (197).

Kennedy y col. (202) describen el algoritmo PSO de la siguiente forma: los individuos (partículas) que conviven en una sociedad tienen una “opinión” que es parte del espacio de búsqueda, compartido por todos los individuos. Cada individuo puede modificar su opinión según tres factores:

- El conocimiento del entorno o “adaptación”.
- Experiencias anteriores del individuo o “memoria del individuo”.
- Experiencias anteriores de los individuos del vecindario o “memoria del vecindario”.

Los individuos adaptan o modifican sus opiniones a las de los individuos con más éxito en su entorno. Con el paso del tiempo, los individuos de un entorno tienen un conjunto de opiniones bastante relacionadas entre ellos.

Las principales características del algoritmo PSO son las siguientes (197):

- En PSO los agentes de búsqueda (partículas) intercambian información. Las partículas modifican su dirección en función de las direcciones de las partículas de su vecindario.
- PSO almacena la experiencia propia o historia de cada agente. La partícula decide su nueva dirección en función de la mejor posición por la que pasó anteriormente.
- Suele tener una convergencia rápida a buenas soluciones.
- La población del algoritmo se inicia de forma aleatoria y evoluciona iteración tras iteración.
- La búsqueda persigue siempre la solución más óptima posible.
- La búsqueda se basa exclusivamente en los valores de la función objetivo.
- PSO trabaja con la información del problema codificada.
- Es una técnica estocástica referida en fases (inicialización y transformación).
- PSO tiene operadores de movimiento pero no de evolución como la mutación o el cruzamiento.

- PSO no crea nuevas partículas durante su ejecución, sino que siempre son las mismas partículas iniciales modificadas a lo largo del proceso.

El pseudocódigo del algoritmo básico de optimización de cúmulo de partículas (PSO) es de la siguiente forma:

Begin

$t = 0$

Nube \leftarrow Inicializar Nube de Partículas

Mientras no se alcance la condición de parada *hacer*

$t = t + 1$

Para $i = 1$ hasta tamaño (Nube) *hacer*

Evaluar cada partícula x_i de la Nube

Si aptitud_ x_i es mejor que aptitud_mejorpos i *entonces*

mejorpos $i \leftarrow x_i$;

aptitud_mejorpos $i \leftarrow$ aptitud_ x_i

Fin Si

Si aptitud_mejorpos es mejor que aptitud_mejorpos *entonces*

mejorpos \leftarrow mejorpos i ;

aptitud_mejorpos \leftarrow aptitud_mejorpos i

Fin Si

Fin Para

Para $i = 1$ hasta tamaño (Nube) *hacer*

Calcular la velocidad v_i de x_i , en base a los valores

x_i , mejor p_{best} y p_{best}

Calcular la nueva posición de x_i , de su valor actual y v_i

Fin Para

Fin Mientras

End (Devuelve la mejor solución encontrada)

Las variantes de este algoritmo básico dependen fundamentalmente de los siguientes factores:

- La implementación de las partículas.
- Cálculo o actualización del vector velocidad.
- Cálculo o actualización del movimiento de las partículas.
- Representación de las soluciones.

Los algoritmos basados en nubes de partículas PSO comparten muchas ideas extraídas de los algoritmos evolutivos, como por ejemplo, la inicialización aleatoria de la población al comienzo del proceso o la búsqueda de soluciones óptimas a través de la mejora generacional de la población. No obstante, los PSO no tienen operadores evolutivos como pueden ser el cruce o la mutación.

Optimización de Colonia de Hormigas (Ant Colony Optimization - ACO)

El algoritmo de optimización basado en la colonia de hormigas o *Ant Colony Optimization* (ACO) (203) y (204) fue introducido por Dorigo en 1992 (203). Es un método probabilístico pensado para resolver problemas combinatorios que está inspirado en el comportamiento que presentan las hormigas para encontrar las trayectorias desde la colonia hasta el alimento. Los estudios realizados explican como animales casi ciegos, como son las hormigas, son capaces de seguir la ruta más corta en su camino de ida y vuelta entre la colonia y una fuente de abastecimiento. Esto es debido a que las hormigas pueden “transmitirse información” entre ellas gracias a que cada una

de ellas, al desplazarse, va dejando un rastro de una sustancia llamada *feromona* a lo largo del camino seguido.

De esta forma, mientras una hormiga aislada se mueve de forma esencialmente aleatoria, los miembros de una colonia de hormigas detectan el rastro de feromona dejado por otras hormigas y tienden a seguir dicho rastro. Éstas, a su vez, van dejando su propia feromona a lo largo del camino recorrido y por tanto lo hacen más atractivo, puesto que se ha reforzado el rastro de feromona. No obstante, la feromona también se evapora con el paso del tiempo provocando que el rastro se difumine y por tanto se debilite. De esta forma, se limita el crecimiento de los rastros por lo que la solución adoptada podría corresponder a un óptimo local.

A modo de resumen, puede decirse que el proceso se caracteriza por una retroalimentación positiva, en la que la probabilidad con la que una hormiga escoge un camino aumenta con el número de hormigas que previamente hayan elegido el mismo camino.

Los algoritmos ACO son procesos iterativos en lo que en cada iteración se “lanza” una colonia de H hormigas y cada una de las hormigas de la colonia construye una solución al problema. Las hormigas construyen las soluciones de manera probabilística, guiándose por un rastro de feromona artificial y por una información calculada *a priori* de manera heurística.

La técnica se basa en dos pasos principales:

- Construcción de una solución basada en el comportamiento de una hormiga.
- Actualización de los rastros de feromona artificiales.

Estos algoritmos de optimización basados en colonias de hormigas - PSO (y otros similares) comparten muchas ideas extraídas de los algoritmos evolutivos, como por ejemplo, la inicialización aleatoria de la población al comienzo del proceso o la búsqueda de soluciones óptimas a través de la mejora generacional de la población. Sin embargo, los PSO no tienen operadores evolutivos como pueden ser el cruce o la mutación.

Se puede obtener más información sobre este algoritmo metaheurístico consultando las fuentes (203) y (204).

El pseudocódigo del algoritmo básico de optimización de colonias de hormigas (ACO) es de la siguiente forma:

Begin

Inicializar parámetros

Mientras condición d

 Inicializar nueva hormiga ()

Mientras estado <> estado fin

Para cada arco posible movimiento

 Calcular probabilidad elección

Fin Para

 Siguiente_posición = política_decisión

 Lista_posiciones = + siguiente_posición

Fin Mientras

 Realizar evaporación

 Depósito feromona (lista_posiciones)

Fin Mientras

End

2 METODOLOGÍA

En esta sección, se describen los materiales y métodos utilizados en esta investigación. En primer lugar, se describe el algoritmo de segmentación propuesto. Para probar sus prestaciones, se realizarán diferentes ejecuciones de él, en diferentes conjuntos de datos. Dichos conjuntos son descritos a continuación, de forma que pueda el lector seguir las diferentes conclusiones que se obtengan en la sección de resultados. Seguidamente, se especificará tanto el software como hardware utilizado. Y por último, se mostrará la encuesta realizada a farmacéuticos de la provincia de Sevilla, como método e instrumento para apoyar la conveniencia de esta investigación.

2.1 El algoritmo de evolución diferencial para segmentación con esquema de continuación (DECCS)

A continuación, mostramos todos los detalles del algoritmo DECCS con idea de que pueda ser implementado por el lector. No obstante, adjuntamos en el Anexo 6.3 el código usado para esta disertación implementado en C++.

2.1.1 Introducción

En esta sección, se explica el funcionamiento del algoritmo de segmentación propuesto. Tal como se ha ido comentando a lo largo de la disertación, nuestro objetivo es diseñar e implementar un algoritmo capaz de resolver problemas de grandes dimensiones. Puesto que actualmente se disponen de innumerables recursos para la obtención de datos, proponemos el uso de algoritmos metaheurísticos (véase la sección 1.4.7.3) que han sido usados con éxito en multitud de problemas en este campo y en otras áreas de investigación. Recordemos que los algoritmos metaheurísticos son algoritmos de optimización aproximados capaces de obtener soluciones, si no óptimas, muy cercanas a las óptimas en tiempos antes inimaginables (más aún si se comparan con los tiempos requeridos por los algoritmos exactos). No obstante, estos algoritmos pueden llegar a ser poco eficientes para problemas de gran tamaño pues requieren de poblaciones de soluciones grandes y por tanto muy costosas de evaluar y evolucionar. Recientemente se ha mostrado la viabilidad del uso de micro poblaciones para abordar problemas de gran tamaño en algoritmos de evolución diferencial mediante la incorporación de

esquemas, poco costosos, que ayuden a mantener la diversidad en estas poblaciones y eviten la convergencia prematura del algoritmo a un óptimo local (véase Segura y col. (207)).

Por todo ello, para el presente trabajo proponemos usar un esquema evolucionado basado en **evolución diferencial** que, como ya se ha comentado anteriormente, sea capaz de resolver, o al menos obtener soluciones de calidad, problemas de segmentación de conjuntos de datos de grandes dimensiones en tiempos reducidos.

A continuación presentamos nuestro algoritmo, que denominamos **DECCS** (del inglés, *Differential Evolution for Clustering with Continuation Scheme*), el cual incorpora las siguientes características que posteriormente detallaremos:

- Codificación de las soluciones con esquema de reordenamiento. Dicha codificación no requiere del conocimiento del número de grupos a buscar sino que determina la mejor solución entre 2 y K_{max} grupos. Además, esta propuesta solventa el clásico problema de simetría de ésta y acelera la convergencia, como se verá en la sección dedicada a los resultados.
- Métodos de muestreo del espacio más inteligentes para generar la población inicial. Como se mostrará, es posible obtener poblaciones iniciales de calidad incluso con tamaños de población reducidos para el esquema de codificación propuesto.
- Incorporación del esquema de continuación propuesto por Segura y col. (207). Mostraremos la validez de dicho esquema para problemas de segmentación de grandes dimensiones. Ello nos permitirá utilizar tamaños de población pequeños, acelerando la convergencia pero sin perder capacidad de exploración.
- Combinación del uso de los operadores de cruce binario y exponencial. Esto lo hace más robusto y flexible para muchos más tipos de problemas.
- Utilización de esquemas auto-adaptativos para sus principales parámetros. Para evitar así la dependencia de sus dos principales parámetros (parámetro de inercia, F , y probabilidad de cruce, CR) y la necesidad de determinarlos previamente, éstos no se consideran fijos. Así, F se seleccionará aleatoriamente

según una distribución uniforme mientras que CR a medida que evoluciona la búsqueda.

2.1.2 Codificación

Proponemos una codificación novedosa para DECCS inspirada en la codificación propuesta por Das y col. (134) para su algoritmo ACDE. En ACDE no se requiere conocer con antelación el número de grupos a determinar pues explora un espacio de soluciones de entre 2 y $Kmax$ grupos, y determina de manera automática el número de grupos más adecuado para el problema. Esto es una gran ventaja, pues no siempre se conoce dicho valor con antelación y no se requiere de un conocimiento previo ni de un análisis previo de éste. Ahora bien, dicha codificación requiere de un número mucho mayor de variables y el problema que se pasa a resolver es aún más complejo. Con este objetivo, se propuso para ACDE una codificación con

$$NVar = Kmax + Kmax * NumCri = Kmax * (NumCri + 1) \text{ variables,}$$

donde $Kmax$ es el número máximo de grupos a determinar y $NumCri$ el número de criterios, o características, del problema.

Veamos un ejemplo de codificación para 2 características y un máximo de 3 grupos. Para estos valores codificaremos cada solución con un vector de $3 \times (2+1) = 9$ variables con la siguiente interpretación:

| | | | | | | | | |
|----|----|----|----|----|----|----|----|----|
| A1 | A2 | A3 | X1 | Y1 | X2 | Y2 | X3 | Y3 |
|----|----|----|----|----|----|----|----|----|

- A_i = variable de activación para el i -ésimo clúster. Toma valores entre 0 y 1 de tal forma que el i -ésimo clúster se activa si A_i es mayor que 0.5 y no se activa en caso contrario.
- (X_i, Y_i) = es el centro del i -ésimo clúster.

Una vez determinado el número de clústeres activos y sus centros, todos los vectores son asignados al grupo cuyo centro sea el más cercano. Una vez asignados, se calculan las distancias intra e inter-grupos según el indicador de calidad utilizado (CS en nuestro caso) y evaluada la solución.

Observación: si el número de clústeres activos es inferior a 2, se activa otro aleatoriamente. Posteriormente, si tras la asignación el tamaño de algún clúster es igual a 1, éste es desactivado (haciendo $A_i = 0$) y reasignado el vector al siguiente clúster más cercano. Si tras este proceso, el número de clústeres resultante es igual a 1, la solución es descartada y se vuelve a generar otra.

Este esquema, presenta un problema de simetría puesto que permutaciones de una misma codificación dan lugar a la misma segmentación. Por este motivo proponemos un esquema de reordenamiento. Dicho esquema consiste en, una vez evaluada la solución y asignados los puntos a los grupos, reordenar las variables colocando los clústeres activos al principio por orden decreciente de tamaño (los empates en tamaños se ordenan al azar).

Veamos un ejemplo. Si la codificación original fue

| | | | | | | | | |
|-----|-----|------|----|----|----|----|----|----|
| 0,3 | 0,7 | 0,55 | X1 | Y1 | X2 | Y2 | X3 | Y3 |
|-----|-----|------|----|----|----|----|----|----|

con dos clústeres activos (el segundo y el tercero pues $A_2 = 0,7$ y $A_3 = 0,55$ ambos mayores que 0,5) de tamaños, por ejemplo $n_2 = 25$ y $n_3 = 43$, la representación de la nueva codificación de dicha solución sería:

| | | | | | | | | |
|------|-----|-----|----|----|----|----|----|----|
| 0,55 | 0,7 | 0,3 | X3 | Y3 | X2 | Y2 | X1 | Y1 |
|------|-----|-----|----|----|----|----|----|----|

Este esquema, salvo para los casos de empate en el tamaño de los grupos, solventa el problema de la simetría. Dejando para la sección 3 los resultados obtenidos de los experimentos realizados para mostrar que el esquema propuesto mejora notablemente el rendimiento del algoritmo ACDE.

Respecto a la complejidad computacional de esta codificación, y dado que los *TamPop* puntos son comparados contra un máximo de K_{max} centros para determinar el más cercano, se requieren de $TamPop * K_{max}$ cálculos de distancias euclídeas, cada una de las cuales depende el número de características o variables del problema ($NumCri$). A modo de ejemplo, para el problema IRIS con $TamPop = 150$, $K_{max} = 6$ y $NumCri = 4$ se realizan $150 * 6 * 4 = 3.600$ operaciones; mientras que para el problema PEN con

$TamPop = 10.992$, $Kmax = 20$ y $NumCri = 16$ se realizan $10.992 * 20 * 16 = 3.517.440$ operaciones (un 877'06% más de tiempo).

El coste computacional del reordenamiento, empleando por ejemplo el algoritmo de la burbuja, es del orden $O(Kmax^2)$. Que es un valor casi despreciable pues $Kmax$ suele ser menor o igual que 20 mientras que $TamPop$ lo supera por mucho.

2.1.3 Evaluación de la calidad. Función objetivo

Como se expuso en la sección 1.4.5, son muchas las métricas o indicadores propuestos en la literatura para cuantificar la calidad de una agrupación que intente reflejar las propiedades deseadas de ésta: homogeneidad en los datos de un mismo clúster a la vez que heterogeneidad entre los grupos. Es más, no existe una métrica única capaz de reflejar con precisión ambos conceptos y suficientemente flexible y versátil como para detectar diferentes geometrías o densidades de los grupos.

En la sección dedicada a los resultados computacionales, mostraremos que la métrica más adecuada para nuestros propósitos es la CS pues presenta las mejores propiedades a la hora de segmentar los datos. Con idea de facilitar la lectura de esta disertación, incluimos la definición de la métrica CS de Chou y col. (149):

$$CS = \frac{\sum_{i=1}^k \frac{1}{N_i} \sum_{x_i \in C_i} \max_{x_j \in C_i} \text{dist}(x_i, x_j)}{\sum_{i=1}^k \min_{j \neq i} \text{dist}(\text{centroide } C_i, \text{centroide } C_j)}$$

El indicador CS compara la suma de la dispersión intra-clúster (numerador) con la inter-clúster (denominador). Por tanto, cuanto menor es el valor obtenido mejor la segmentación obtenida. Es más eficiente que otros indicadores con clúster de distinta densidad, aunque su coste computacional es más elevado.

Los resultados experimentales que avalan nuestra selección van en concordancia con los obtenidos por Das y col. (134) para su algoritmo ACDE y se han incluido en la sección 3 para facilitar la exposición del algoritmo. No obstante, como se expondrá también dicha sección, si se desea menor sensibilidad a formar grupos y obtener un número menor de clústeres, la métrica de *Calinsky-Harabasz* es una buena opción.

2.1.4 Población inicial

Para generar la población inicial del algoritmo DECCS proponemos usar el método basado en **hipercubos latinos**. Dicho método ha sido exitosamente usado en estudios previos (véanse (208), (209) y (210)). En dichos trabajos se exponen cómo la generación de soluciones de manera controlada pero con cierta componente estocástica mejora las propiedades de convergencia de los algoritmos desde dos puntos de vista: tanto en la aceleración de la convergencia como en la calidad de las soluciones finales obtenidas.

El método de muestreo basado en hipercubos latinos (*Latin Hypercubes Sampling*) es un método estadístico propuesto para generar una muestra de posibles conjuntos de valores de los parámetros de una distribución multidimensional. El método fue descrito originalmente por McKay en 1979 (211) aunque proponemos la lectura de Wyss y Jorgensen (212) para profundizar en los métodos y sus variantes.

En el contexto de muestreo estadístico, una cuadrícula que contiene posiciones de muestra es un cuadrado latino si (y sólo si) hay sólo una muestra en cada fila y cada columna. Un hipercubo latino es la generalización de este concepto a un número arbitrario de dimensiones, por lo que cada muestra es la única en cada hiperplano que lo contiene.

En nuestro caso, si el problema depende de $NVar$ variables y deseamos un tamaño de población de $TamPop$ puntos, el rango de cada variable se divide en $TamPop$ intervalos igualmente probables. De esta forma el espacio de variables queda dividido en $TamPopNVar$ hipercubos. Posteriormente, $TamPop$ puntos se colocan en el hipercubo satisfaciendo los requisitos de un hipercubo latino: debe haber exactamente un punto en cada fila y en cada columna de la hipermatriz.

Veamos un ejemplo sencillo que ilustra perfectamente la metodología. Supongamos $Nvar = 2$ con $x_i \in [-1,1]$ para ambas variables y $TamPop = 4$. Una muestra posible es la contenida en la Figura 78. En dicha figura se aprecia como el muestreo realizado garantiza exactamente una solución en cada uno de los 4 subintervalos de ambas variables.

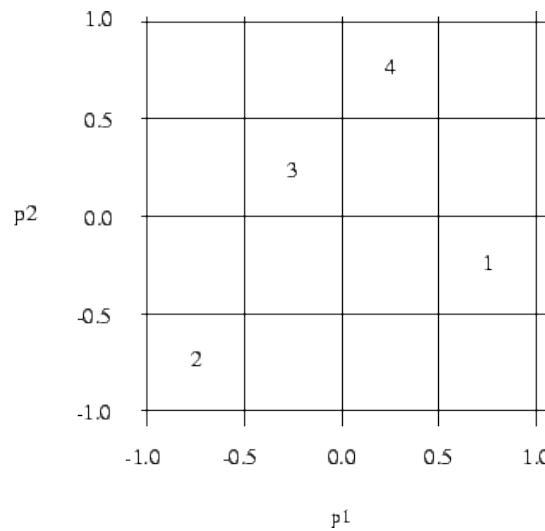


Figura 78. Muestra de tamaño 4 mediante Hipercubos latinos para un espacio bidimensional

En la sección 3, dedicada a los resultados experimentales, comparamos este método frente a dos procedimientos alternativos, vectores opuestos y muestreo aleatorio, sobre cuatro problemas de segmentación. El objetivo de estos experimentos es doble. Por un lado, validar la selección realizada y por otro determinar el tamaño de la población más adecuado para DECCS.

2.1.5 Mutación y cruce

Los operadores de mutación y cruce utilizados en DECCS son los ya explicados en la sección 2 con las siguientes modificaciones:

- Los operadores de cruce binario y exponencial se usan indistintamente. Se genera un valor aleatorio siguiendo una distribución $U(0,1)$. Si dicho valor es menor que 0,5 se aplica el binomial, en caso contrario el exponencial. Ambos operadores incorporan el esquema de continuación propuesto por Segura y col. (207), que potenciarán la capacidad de búsqueda de DECCS a pesar de usar tamaños de población pequeños. Se expondrá con más detalle en la siguiente sección.
- El parámetro CR o probabilidad de cruce no es fijo y evoluciona con la búsqueda tratando de reforzar las fases de exploración y explotación de DECCS. Así, el

parámetro CR decrece linealmente de 1 a 0.5 en las *NumIter* iteraciones del algoritmo, cuya expresión es la siguiente:

$$CR = 0,5 + 0,5 * \frac{NumIter - iter}{NumIter}$$

siendo *iter* la iteración actual y *NumIter* el número máximo de iteraciones a realizar.

- El factor de inercia, *F*, es seleccionado aleatoriamente siguiendo una distribución uniforme $U(0,5; 1,5)$. De esta forma, se evita tener que ajustarlo con antelación y que los resultados de DECCS puedan depender de dicho valor en gran medida.

2.1.6 El Esquema de continuación

Segura y col. (2015) analizan dos de los problemas clásicos en evolución diferencial para problemas en conjunto de datos de alta dimensionalidad. Estos son: la pérdida de diversidad en la población y la rápida reducción de capacidad exploradora de los operadores de cruce y mutación. Los autores proponen dos esquemas para solventarlos y lo prueban en dos conjuntos clásicos de problemas de gran tamaño. En la presente disertación se mostrará la validez de éstos para problemas de segmentación codificados tal y como se expuso en la sección 2.1.2.

Paseamos a detallar los dos esquemas incorporados a DECCS.

2.1.6.1 Continuación. Falta de diversidad en los vectores diferencia

El objetivo de este primer esquema es incrementar el número de potenciales diferencias que pueden ser generadas con una evolución diferencial. Nuestra propuesta está basada en la generación de funciones continuas acumulativas no decrecientes (CDF) en el rango de posibles diferencias a partir de todas las posibles diferencias, en valor absoluto, entre dos elementos cualesquiera de la población. Entonces, la CDF generada es usada para generar números aleatorios que son multiplicados luego por *F* y usados para mutar a los vectores de la población.

Esto se hace a partir de la ecuación siguiente:

$$CDF(x) = \frac{minor(x)}{0.5 * TamPop * (TamPop - 1) - 1}$$

donde $minor(x)$ es el número de diferencias que aparecen en la población menores que x . El denominador es la cantidad de vectores diferencia que aparece en la población menos 1. Esto nos asegura que el máximo valor de $CDF(x)$ sea 1. La función $CDF(x)$ tiene forma de escalera. Posteriormente, dicha función es “suavizada” interpolando linealmente entre cada par de valores consecutivos. Así, obtenemos una función continua con probabilidad no nula para todas las diferencias entre la mínima y la máxima.

Veamos un ejemplo. Supongamos una población de tamaño 5, $TamPop = 5$, con las siguientes 10 diferencias (en términos absolutos): 1,1 1,3 1,4 1,41, 1,43 1,45 8, 8,3 8,5 y 8,7.

La Figura 79 muestra la función $CDF(x)$ con estos 10 valores en el rango $[1,1; 8,7]$.

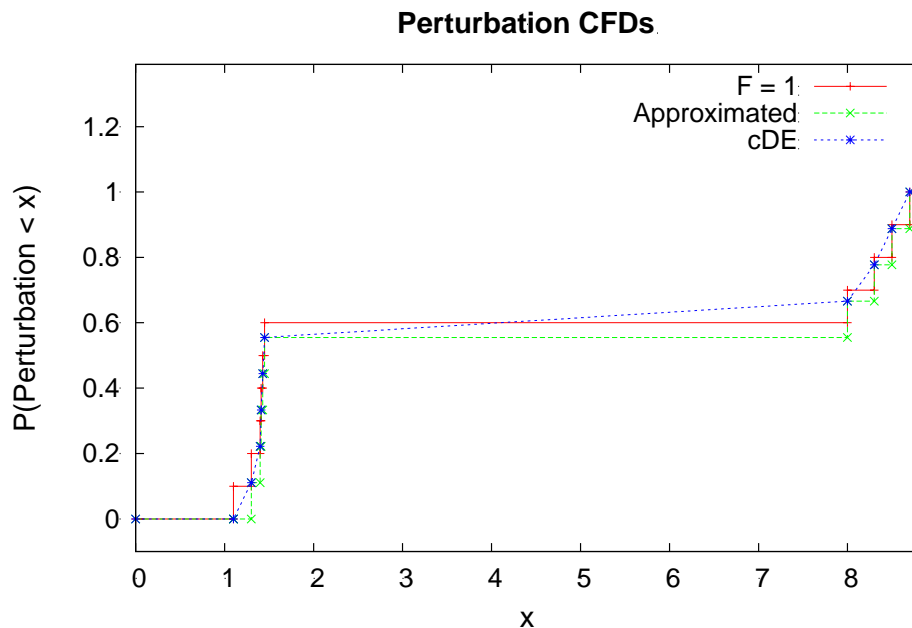


Figura 79. Ejemplo de función $CDF(x)$

A este esquema es al que denominamos esquema de continuación pues cualquier valor de mutación entre el mínimo y el máximo, $[1,1; 8,7]$ en el ejemplo, puede ser generado. No obstante, es importante resaltar que *CDF* tiene pendientes mayores en aquellas zonas con más acumulación de valores o diferencias. Así, no se pierden los principios que subyacen en una evolución diferencial a la vez que aumentamos el conjunto de posibles mutaciones, evitando la convergencia prematura a un óptimo local.

2.1.6.2 Perturbaciones grandes

El uso ocasional de mutaciones grandes mediante la generación de valores grandes de *F*, ha resultado beneficioso en ciertos problemas pues permite escapar de los óptimos locales. No obstante, también se generan multitud de soluciones no factibles que no resultan deseables. Por ello, el siguiente esquema promueve el uso esporádico de grandes mutaciones pero de manera controlada.

A lo anterior debemos añadir que la probabilidad de que las mutaciones, o vectores diferencias, sean pequeñas aumenta con el número de variables. Esto puede ocurrir incluso en las primeras iteraciones de la búsqueda. De hecho, notemos que dados dos vectores o soluciones cualesquiera (X_i, X_j) , la probabilidad de que la diferencia entre los valores de su correspondiente variable *v*-ésima sea menor que ε es:

$$P(|X_{i,v} - X_{j,v}| < \varepsilon) = \frac{2\varepsilon}{cotasup_v - cotainf_v} - \frac{\varepsilon^2}{(cotasup_v - cotainf_v)^2}$$

Obviamente, a medida que el número de variables aumenta, la probabilidad de que estas diferencias sean pequeñas, en al menos una variable, aumenta.

Para evitar esta pérdida de diversidad inherente a los esquemas basados en evolución diferencial, en especial con poblaciones pequeñas, se introduce un esquema adaptativo que promueve perturbaciones largas con probabilidades asociadas bajas. Más concretamente, cada vez que el valor *L* generado por el cruce exponencial es igual a 1, este nuevo esquema se usa con probabilidad *HMR* (del inglés, *High Mutation Ratio*). En los restantes casos (probabilidad $1-HMR$), se aplica el esquema de continuación

anteriormente descrito. Nótese que usar perturbaciones grandes con valores grandes de L podría degenerar en una búsqueda demasiado disruptiva.

El esquema almacena la máxima permutación admisible ($FalseMax_i$) para cada variable i . Inicialmente, son usados los siguientes valores:

$$FalseMax_i = \frac{cotasup_i - cotainf_i}{5}$$

Así, inicialmente se permiten perturbaciones máximas de hasta el 20% del rango de la variable. Cada vez que se usa este mecanismo para, por ejemplo, la variable i el operador actúa como sigue. En primer lugar, se calcula la perturbación máxima que puede ser generada con los vectores diferencia de la población (véase la definición de $CDF(x)$ en la sección anterior), que denotamos Max_i . Posteriormente, se genera un número aleatorio en el intervalo $[Max_i, FalseMax_i]$ mediante una distribución uniforme. Este número es usado para perturbar el vector base. En la mitad de los casos se usa con signo positivo mientras que en el resto se usa con signo negativo. Finalmente, el valor de $FalseMax_i$ es actualizado. En los casos en los que el vector generado mejora al vector objetivo, el valor se reajusta con:

$$FalseMax_i = FalseMax_i + \frac{FalseMax_i - Max_i}{10}$$

Mientras que en los casos en los que no se obtienen mejoras se opta por decrecerlo de la siguiente forma:

$$FalseMax_i = FalseMax_i - \frac{FalseMax_i - Max_i}{10}$$

El principio que subyace bajo este esquema de adaptación, es promover perturbaciones incluso más grandes cuando han sido exitosamente utilizadas y reducirlas cuando no.

Obviamente, estos valores utilizados para Max_i y $FalseMax_i$ pueden ser seleccionados de otra forma según interese. Si se desean perturbaciones aún más grandes, basta con modificar el denominador de $FalseMax_i$ o adaptar sus valores a mayor o menor

velocidad.

2.1.7 Pseudocódigo de DECCS

A continuación mostramos el pseudocódigo de DECCS con las principales fases que la componen. Recordamos no obstante, que el código completo en C++ se adjunta en el Anexo 6.3.

En primer lugar, se genera la población inicial de tamaño *TamPop* con el método de muestreo basado en hipercubos latinos. Posteriormente, la población evoluciona un número de iteraciones igual a *NumIter*. En cada iteración, y siguiendo un esquema basado en evolución diferencial, cada elemento de la población es considerado vector objetivo y trata de ser reemplazado por uno mejor generado por los operadores de mutación y cruce, binario y exponencial, propios de una evolución diferencial. En dichos operadores se incorporan el esquema de continuación (tanto para binario como exponencial) y el esquema de perturbación (solo para el esquema exponencial y con probabilidad HMR). Si la solución generada mejora al vector objetivo, lo reemplaza. Si no, se rechaza. A su vez, si el vector generado mejora al óptimo encontrado hasta el momento, la reemplaza. El número total de evaluaciones realizadas de la función objetivo no es más que el producto de *TamPop* y *NumIter*. Puesto que ACDE y DECCS usan tamaños diferentes, hemos usado el valor de *NumIter* para conseguir que ambos algoritmos realicen el mismo número de exploraciones y que la comparación sea más justa entre ellos.

Algoritmo. DECCS (Differential Evolution for Clustering with Continuation Scheme)

Input: *TamPop*, *NumIter*, *Kmax*, *HMR*

Output: segmentación óptima (1er elemento de la población, *Pop* [0])

Pop[*i*] ← GenerarPopInicial(); /*Población inicial de tamaño *TamPop* con Hipercubos Latinos*/

/*Guardamos el óptimo en *Pop*[0]*/

FOR iter = 1...*NumIter*

FOR i = 1...*TamPop*

Seleccionar 3 elementos distintos de la población

IF ($U(0,1) < 0,5$) /*Se utilizan BINARIO y EXPONENCIAL lanzando una moneda*/

Sol ← Operador de cruce BINARIO con esquema clásico y con esquema de continuación en una variable

ELSEIF

IF ($U(0,1) < HMR$)

Sol ← Operador de cruce EXPONENCIAL con esquema clásico y con esquema de continuación (perturbación) en 1 sola variable

ELSE

Sol ← Operador de cruce EXPONENCIAL con esquema de continuación en sola 1 variable

ENDIF

ENDIF

Sol ← Comprobar factibilidad () /*Comprobar si es factible y arreglar si no*/

Sol ← Evaluar ()

IF (*Sol*.f < *Pop*[*i*].f)

Pop[*i*] = *Sol*

IF *Sol*.f < *Pop*[0].f

Pop[0] = *Sol*

ENDIF

ENDIF

ENDFOR

ENDFOR

2.2 Benchmark en segmentación

Son muchas y muy variadas las bases de datos con problemas de segmentación. Estas bases de datos se pueden clasificar de muchas maneras. A continuación enumeramos algunas posibles subdivisiones:

- Problemas reales o sintéticos. En función de los objetivos establecidos en la investigación pueden interesar unos u otros. En ocasiones ocurre que no encontramos problemas reales con ciertas características (nivel de solape entre clústeres, densidades o variabilidades de clústeres muy descompensados, geometrías y formas complicadas de detectar por un algoritmo convencional, etc.). En estas ocasiones, es habitual que los investigadores diseñen *ad hoc* sus propios problemas con los que testar los algoritmos y comparar sus habilidades para manejar y solventar situaciones atípicas.
- Naturaleza de las variables. Según las variables sean de naturaleza real, entera, nominal o una combinación de éstas.
- Tamaño de la muestra. Este valor indica el tamaño de puntos o vectores a analizar. Es cada vez más habitual encontrarse problemas de gran escala que han dado origen a lo que hoy en día se llama *Big Data*.
- Número de clases (si se conocen). En los problemas de segmentación no es habitual conocer el número de clústeres a priori pero sí en los de clasificación. Cuando se desea probar la habilidad de un algoritmo para detectar el número óptimo de clúster es habitual recurrir a problemas de clasificación para tener un valor contra el que compararse.
- Otros. En este grupo incluimos aquellos problemas incompletos o con valores perdidos así como los que incluyen algo de “ruido”.

La mayoría de los problemas están disponibles en distintos sitios web con bases de datos no solo para segmentación sino también para problemas de clasificación, series temporales, decisión, etc. A continuación enumeramos las más populares siendo la primera de ellas (*UCI Machine Learning Repository*) una de la más citadas (147).

- **UCI Machine Learning Repository** (147). Se actualiza periódicamente contando en la actualidad con 234 problemas de clasificación, 53 de regresión, 42 bases de datos para problemas de segmentación y 50 de otro tipo. Además, se pueden clasificar según sean univariantes o multivariantes, de series temporales, secuenciales o de textos.
- **KEEL-data set** (Knowledge Extraction based on Evolutionary Learning) (148). Creada por un grupo de la Universidad de Granada, contiene además de multitud de bases de datos, un software *ad hoc* para visualizarlas, editarlas, y desarrollar experimentos.
- **Speech and Image Processing Unit** (149). Web más modesta con una mezcla de bases de datos tanto de procesamiento de imágenes como de bases de datos artificiales.
- **Data sets for Clustering Techniques de la Universidad de Colonia** (150). Contiene 5 problemas reales de tamaño medio (con entre 2 y 7 variables).
- **CIFAR-10 y CIFAR-100** (151) para reconocimiento de imágenes. CIFAR-10 contiene millones de imágenes clasificables en 10 grupos con 6000 imágenes cada uno, mientras que CIFAR-100 clasifica las mismas imágenes en 100 grupos con 600 imágenes cada una.

Los problemas que se han seleccionado para validar nuestro algoritmo los hemos clasificado según sean de tamaño medio o de gran tamaño. El algoritmo que proponemos en los próximos capítulos incorporará mecanismos específicamente diseñados para bases de datos de gran tamaño. Por este motivo se han seleccionado varios problemas con un gran volumen de datos y/o variables sobre los que comparar el rendimiento del algoritmo con y sin dichos mecanismos. Como se explicará en el próximo capítulo, el número de variables necesarias para codificar el problema es igual a:

$$NCluster + NCluster \times NVariables = NCluster \times (NVariables + 1)$$

siendo *NCluster* el número máximo de clústeres permitidos y *NVariables* el número de atributos o variables que contiene. Además, el coste computacional del algoritmo depende obviamente del número de datos o puntos que contienen.

2.2.1 Problemas de tamaño medio

En esta sección, incluimos las bases de datos seleccionadas para esta categoría. Se han seleccionado los problemas más clásicos puesto que esto ayuda a la hora de realizar comparaciones con otros algoritmos. Son problemas bien conocidos también usados en Das, Abraham y Konar (134) contra los que nos compararemos posteriormente.

- **IRIS** (4 variables, 3 grupos y 150 vectores). Esta es quizás la base de datos más conocida de la literatura de reconocimiento de patrones. El conjunto de datos contiene 3 clases de 50 casos cada uno, donde cada clase se refiere a un tipo de planta Iris. Una clase es linealmente separable de las otras dos pero estos dos grupos no son linealmente separables uno del otro.



Figura 80. IRIS

- **GLASS** (9 variables, 6 grupos y 214 vectores) Propuesto por el Servicio de Ciencias Forenses de los EE.UU, contiene 6 tipos de cristal que pueden ser encontrados en la escena de un crimen y definidos en función del contenido de óxido (i.e., Na, Fe, K, etc.).



Figura 81. GLASS

- **WINE** (13 variables, 3 grupos y 178 vectores). Estos datos son el resultado de un análisis químico de los vinos cultivados en una misma región de Italia, pero derivados de tres cultivos diferentes. El análisis determina las cantidades de 13 componentes que se encuentran en cada uno de los tres tipos de vinos. En un contexto de clasificación, este es un problema bien planteado con estructuras claramente diferenciadas por lo que no es muy difícil de resolver.



Figura 82. Wine

- **WISCONSIN BREAST CANCER** (9 variables, 2 grupos y 683 vectores). Este problema contiene los casos de un estudio que se realizó en los hospitales de la Universidad de Wisconsin (Madison) sobre los pacientes que se sometieron a cirugía al sufrir de cáncer de mama. La tarea consiste en determinar si el tumor es benigno o maligno.



Figura 83. Winsconsin Breast cancer

- **THYROID** (5 variables, 2 grupos y 215 vectores). Este conjunto de datos es una de las varias bases de datos sobre la tiroides disponible en el repositorio de UCI. La tarea es detectar si un paciente dado es normal o sufre de hipertiroidismo o hipotiroidismo.



Figura 84. Thyroid

- **S-SET** (2 variables, 15 grupos y 5000 vectores). Son 4 problemas bidimensionales no reales con 5000 vectores y 15 clústeres generados a partir de distribuciones gaussianas con diferente grado de solape.

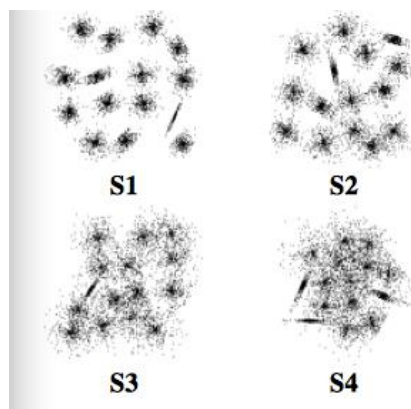


Figura 85. S-Set

En resumen, se han seleccionado problemas que requerirán entre 24 y 120 variables, con entre 2 y 13 variables y hasta 5000 vectores. No son problemas fáciles como veremos a continuación, aunque obviamente se pueden encontrar de mayor tamaño. En la próxima sección, seleccionaremos otros aún más complejos.

2.2.2 Problemas de gran tamaño

- **LED Display Domain data set** (7 variables, 10 grupos y 500 vectores). Este problema contiene 7 atributos binarios, uno para cada diodo emisor de luz de un *display* de 7 segmentos. La tarea es determinar qué dígito se ha mostrado en la pantalla. El problema sería fácil si no fuese por la introducción de “ruido” pues cada valor del atributo tiene el 10% de probabilidad de tener su valor invertido. La muestra contiene 500 casos obtenidos a partir de los datos originales.

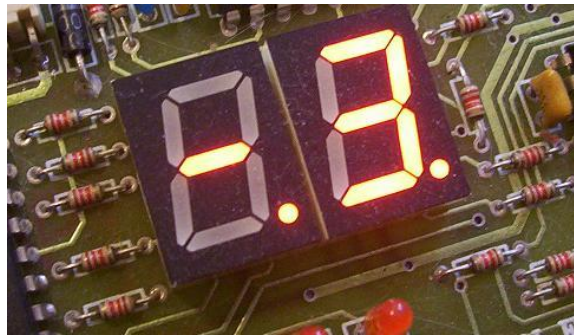


Figura 86. LED Display Domain data set

- **YEAST** (8 variables, 10 grupos y 1484 vectores). Esta base de datos contiene información sobre un conjunto de células de levadura. La tarea consiste en determinar la localización de cada celda entre 10 alternativas posibles a partir de las siguiente variables:
 - Mcg: el método de McGeoch de reconocimiento de secuencias de señales.
 - GvH: método de von Heijne de reconocimiento de secuencias de señales.
 - Alm: Puntuación del programa de predicción de la región que abarca la membrana de ALOM.
 - Mit: Puntuación del Análisis Discriminante del contenido de aminoácidos de la región N-terminal (20 residuos de longitud) de proteínas mitocondriales y no mitocondriales.

- Erl: Presencia de la cadena "HDEL" (pensado para actuar como una señal para la retención en el lumen del retículo endoplásmico). Es por tanto un atributo binario.
- Viruela: Señal de Peroxisomal en el término C.
- Vac: Puntuación del análisis discriminante del contenido de aminoácidos de vacuolar y proteínas extracelulares.
- Nuc: Puntuación del análisis discriminante de las señales de localización nuclear de proteínas nucleares y no nucleares.

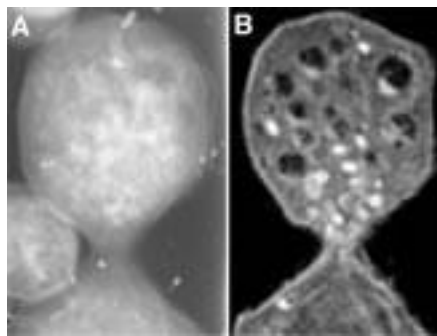


Figura 87. Yeast

- **Red Wine Quality data set** (11 variables, 11 grupos y 1599 vectores). Conjunto de datos relacionado con la variante de color rojo del vino portugués Vinho Verde. Debido a cuestiones de privacidad y logísticos, únicamente las variables fisicoquímicas (*inputs*) y sensoriales (*outputs*) están disponibles (por ejemplo, no hay datos sobre los tipos de uva, marca de vino, precio de venta de vino, etc.). Las clases están ordenadas y no balanceadas (por ejemplo, hay vinos más “normales” que excelentes o de baja calidad).



Figura 88. Red Wine Quality data set

- ***Pen-Based Recognition of Handwritten Digits data set*** (16 variables, 10 grupos y 10992 vectores). Es una base de datos que recopila 250 muestras de 44 escritores. A partir de 16 variables seleccionadas se trata de reconocer el dígito, del 0 al 9, escrito.



Figura 89. Pen-Based Recognition of Handwritten Digits data set

- ***Sonar, Mines vs Rocks data set*** (60 variables, 2 grupos y 208 vectores). Este problema contiene señales emitidas por un sensor desde una variedad de ángulos diferentes, que abarca 90 grados para las minas y 180 grados para las rocas. Cada patrón es un conjunto de 60 números en el rango de 0.0 a 1.0, donde cada número representa la energía dentro de una banda de frecuencia particular, integrada a lo largo de un cierto período de tiempo. Se trata de clasificarlas correctamente según sea una mina o una roca.



Figura 90. Sonar, Mines vs Rocks data set

2.2.3 Conjunto de datos de una oficina de farmacia

A la hora de ejecutar el algoritmo propuesto en esta investigación, para su aplicación en una oficina de farmacia, se ha elaborado un conjunto de datos con información sobre las ventas a clientes en una oficina de farmacia. Concretamente en la farmacia de Manuel Ojeda; una farmacia urbana de tipo medio situada en un barrio joven de Sevilla.

Dicha farmacia, ha evolucionado desde su ubicación inicial en un barrio cuyos pacientes tenían una edad media alta, hasta su nueva ubicación en un barrio joven y con proyección de crecimiento en el que la edad de su población de referencia ha descendido de forma notoria con respecto a la de su ubicación anterior. Todo ello ha provocado la evolución de la oficina de farmacia, que además incluye óptica, y la consideración de los términos “pacientes” y “clientes”, lo que justifica su elección como oficina de farmacia en la que ejecutar nuestros experimentos.

En esta farmacia se distinguen oportunamente los conceptos de cliente y paciente, tal y como se hacen referencia en esta investigación. Para ello, la farmacia ha creado la llamada “*tarjeta saludable*” que permite la obtención de puntos al adquirir productos de parafarmacia, nunca por compra de medicamentos, que posteriormente puede canjear por euros para la realización de nuevas compras de parafarmacia. En concreto, la gestión de clientes del programa “*tarjeta saludable*”, la realiza con un aplicativo independiente de su programa de gestión que recoge la información que se detalla en el Anexo 6.1. Dicho aplicativo, realizado por el autor de esta investigación, permite el registro de toda la información de los clientes y su posterior explotación.

En el momento de realizar esta investigación, se está desarrollando la integración entre el programa de gestión y el aplicativo de gestión de pacientes, por lo que su análisis conjunto no es posible. No obstante, se ha conformado una base de datos aún para esta investigación, conformada por 6 variables y 13.542 registros. Los datos se corresponden con las ventas realizadas en la farmacia durante todo el año 2014. Es importante destacar que, para el acceso a esta información, el autor de esta investigación ha firmado un acuerdo de confidencialidad y no divulgación de los datos a nivel individual. Además, los datos han sido tipificados, con su media y su desviación típica, que son lo que se han manejado, que junto a la no identificación a nivel individual, lo hace más seguro.

Dicha base de datos, depurada previamente, conforma un conjunto de datos sobre el que ejecutar el algoritmo de evolución diferencial mejorado que se propone en este trabajo.

Las características de las variables de este conjunto de datos aparecen en la Tabla 48:

| Variable | Descripción |
|---------------------|--|
| PVP_VL | Importe del ticket de la venta realizada para medicamentos OTC |
| Unidades_VL | Unidades incluidas en el ticket de la venta realizada para medicamentos OTC |
| PVP_SOE | Importe del ticket de la venta realizada para medicamentos con receta |
| Unidades_SOE | Unidades incluidas en el ticket de la venta realizada para medicamentos con receta |

Tabla 48. Características de las variables del conjunto de datos de la oficina de farmacia

2.3 Software y hardware utilizados

Toda la experimentación se ha realizado en un equipo portátil *Apple MacBook Pro* con procesador 2.2 GHz *Intel Core i7* y 16 GB de memoria.

En cuanto al software utilizado, se ha trabajado tanto con software comercial como no comercial.

Todas las ejecuciones realizadas del algoritmo de segmentación propuesto, junto con ACDE, se han producido con un software propio desarrollado en el marco de esta investigación. Para ello, se ha utilizado C++. Que es un lenguaje de programación orientado a objetos. El código de dicho software se detalla en la sección 6.3.

Para obtener los resultados con los algoritmos *K-means*, bietápico y jerárquico, que se comparan tanto con ACDE como DECCS, se ha utilizado el software comercial *IBM SPSS 22*.

Para diseñar el cuestionario de las oficinas de farmacia, que se expone en el Anexo 6.3, y recoger sus datos, se ha utilizado Google Docs. Para el tratamiento y análisis de ha utilizado MS Excel y *IBM SPSS 22*.

2.4 Encuesta a farmacéuticos de la provincia de Sevilla

Aunque, como se ha visto en la sección 1.2, la situación actual de la oficina de farmacia ha activado a la comunidad de farmacéuticos, se ha considerado relevante en esta investigación conocer, a través de un contacto directo con ellos, si el concepto de “cliente” y el establecimiento de la relación con él, está presente en las estrategias de las oficinas de farmacia de la provincia de Sevilla.

Así, se ha realizado un estudio con motivo de conocer el grado de adecuación del problema planteado en esta investigación, con la realidad de las oficinas de farmacia sevillanas.

Para ello, se ha optado por diseñar un cuestionario, que se ofrece en el Anexo 6.3, para ser cumplimentado por diferentes oficinas de farmacia en la provincia de Sevilla con la colaboración del Real e Ilustre Colegio de Farmacéuticos de Sevilla. Con dicho motivo, se envió una circular a todos los colegiados con la petición de cumplimentar la encuesta cuyo link se adjuntaba en la propia circular. La encuesta, se colocó en la plataforma Google Docs lo que facilitó su cumplimentación, depuración y recogida de datos.

Aunque el conocimiento de esta realidad no es propiamente el objeto de esta investigación, se ha considerado interesante pulsar la opinión del farmacéutico de a pie para calibrar la adecuación de la investigación realizada. Dado que no es objeto propiamente dicho de esta investigación, se ha procedido a enviar a todos los farmacéuticos de Sevilla el link del cuestionario para que de forma anónima y voluntaria fuese cumplimentado. De esta forma, también queremos conocer el nivel de respuesta y éste puede ser tratado como un indicador de interés sobre el tema planteado.

El envío tuvo lugar en el mes de diciembre de 2014 y se cumplimentaron 183 cuestionarios por parte de 183 oficinas de farmacia. El trabajo de campo duró una semana. Los principales resultados se observan en la sección 3.6.

3 RESULTADOS Y DISCUSIÓN

3.1 Introducción

En este capítulo, presentamos los resultados obtenidos con el objetivo de validar tanto el algoritmo propuesto, denominado *DECCS*, como las decisiones tomadas durante su diseño e implementación. Más concretamente, decisiones como la selección de la función objetivo, que cuantifique la calidad de las soluciones, la codificación propuesta para solventar el problema de la simetría o el uso de métodos de muestreo más inteligentes para generar tanto la población inicial como el tamaño de ésta.

En la mayoría de los casos, y siempre que los tiempos de cómputo lo permitan, se acompañarán los resultados con un análisis estadístico de los resultados. Así, cada problema se resolverá 40 veces con cada uno de los algoritmos. De esta forma, realizaremos el correspondiente contraste *U-Mann-Withney* para contrastar significativamente la diferencia de medias entre dos algoritmos o bien se realizará un contraste de *Kruskal-Wallis* para más de dos algoritmos. Aunque el nivel de significación habitual fijado es del 5%, veremos como los p-valores obtenidos son o bien 1, concluyendo que no existen diferencias significativas entre los algoritmos comparados, o bien valores muy próximos a 0, con lo que concluiremos la superioridad evidente de un método frente a otro.

Así, los resultados obtenidos que aparecen en el presente capítulo se dividen en los siguientes grandes bloques:

- **Bloque 1.** Comparativa entre cuatro indicadores de validez de la segmentación de gran uso en la literatura para problemas de segmentación. Esta comparación se expone en la sección 3.2.
- **Bloque 2.** Comparativa de resultados entre un algoritmo de segmentación clásico (ACDE) con y sin esquema de reordenamiento de las variables para solventar el problema de la simetría en la codificación. Estos resultados se muestran en la sección 0.
- **Bloque 3.** Comparativa de tres métodos de muestreo del espacio de variables sobre problemas de segmentación para la codificación con reordenamiento. Esta comparación se muestra en la sección 3.4.

- **Bloque 4.** Comparativa final de los resultados obtenidos con DECCS para problemas de tamaño medio y de gran tamaño. Los resultados se compararán contra ACDE y otros tres métodos clásicos: *K-means*, jerárquico y bietápico. Además, los resultados se acompañarán de una comparativa de velocidades y tiempos de convergencia. Finalmente, se mostrarán los resultados obtenidos con DECCS sobre una aplicación real a una oficina de farmacia. Este bloque final se muestra en la sección 3.5.

3.2 La función objetivo. Indicadores de calidad

En este primer bloque de resultados experimentales, mostramos los resultados obtenidos al comparar cuatro de las principales métricas o indicadores de validez de una segmentación. Es bien conocido, y ya se ha expuesto con anterioridad en esta disertación, que no existe un solo indicador capaz de medir o cuantificar con precisión la “calidad” de una segmentación que sea válida para cualquier problema de segmentación. Hay que tener en cuenta que son muchos los aspectos a tener en cuenta: número de grupos, diversidad de tamaños de éstos, diferentes geometrías, diferentes nociones de compacidad u homogeneidad de los grupos, etc., que ya se trataron en la sección 1.4.5.

Por todo lo anterior, hemos seleccionado cuatro de las principales métricas cuyas definiciones pueden consultarse en dicha sección. Son las métricas *CS*, *DN*, *CH* y *DB*.

A partir del conocimiento de los grupos óptimos de varios problemas es posible seleccionar el indicador que mayor habilidad posea para detectarlos y clasificar los datos correctamente según lo previsto.

Conclusión

Como se expondrá a continuación, proponemos el uso de dos métricas según los objetivos que se pretendan alcanzar:

- La métrica CS si deseamos obtener un mayor número de grupos o perfiles de usuarios, pues se muestra más sensible a la hora de crear más grupos de menor tamaño.
- La métrica CH si desea un menor número de grupos de mayor tamaño.

Para llegar a la anterior conclusión, se ha procedido de la siguiente forma. Se han seleccionado los problemas *S1*, *S2* y *S3* (véase la sección 2.2.1). Estos problemas son similares en el número de variables, grupos y volumen de datos pero, como se ha visto con anterioridad, se diferencian claramente en la nitidez o nivel de solape entre los grupos. Así, *S1* es el problema cuya separación entre los grupos es más evidente, en *S2* la separación es menor, mientras que *S3* presenta solapes de gran dificultad a la hora de la segmentación. Puesto que se conoce la distribución “óptima” de éstos, es posible evaluarlas con las cuatro métricas seleccionadas y cuantificar tanto la distancia al valor óptimo como si existen mejores particiones según la métrica.

Por otro lado, puesto que no es necesario que trabajemos con quince grupos para determinar la mejor de las métricas, y con la idea de simplificar la experimentación, los tres problemas han sido truncados para quedarnos únicamente con 3, 5 y 7 grupos de los quince. Así, el problema *S1* tendría los siguientes óptimos para 3, 5 y 7 grupos:

- Para 3 grupos, los grupos son muy claros aunque la geometría lineal del clúster 2 lo diferencia de los otros dos clústeres (Figura 91).
- Para 5 grupos existe un pequeño solape entre los grupos 2 y 5 (Figura 92).

- Mientras que para 7 grupos se añade otro clúster lineal (clúster 6) y otro cercano a éste (clúster 7) como se observa en la Figura 93.

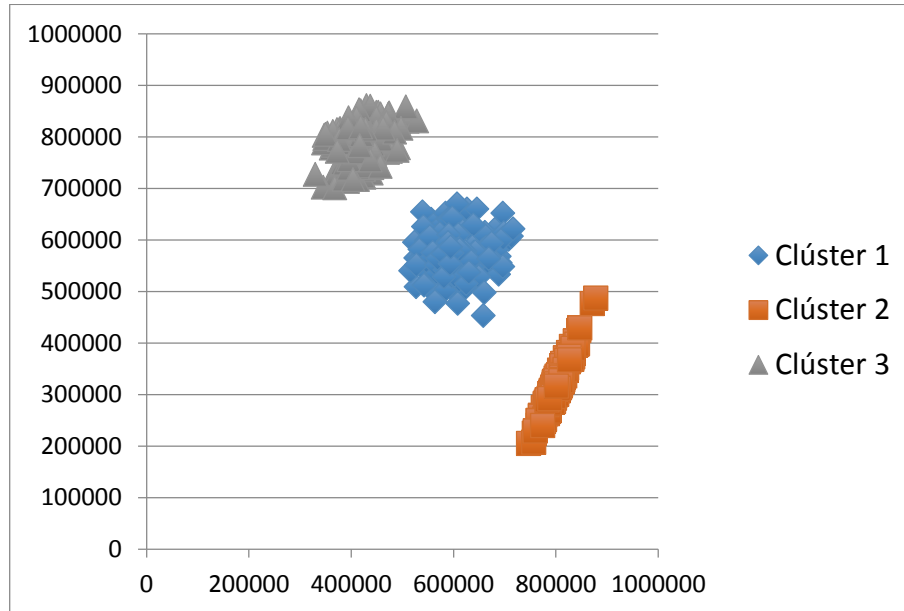


Figura 91. Conglomerados óptimos para el problema S1 truncado a 3 grupos

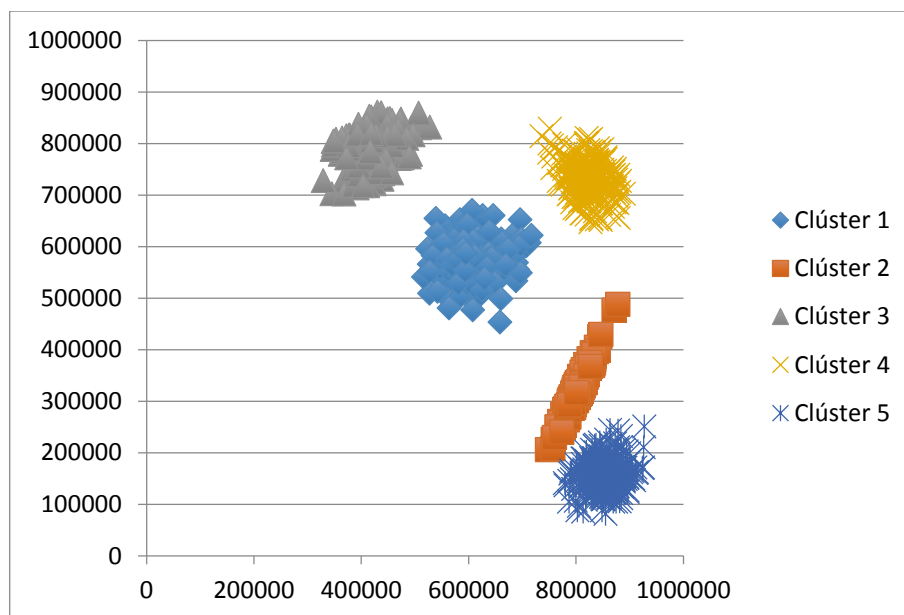


Figura 92. Conglomerados óptimos para el problema S1 truncado a 5 grupos

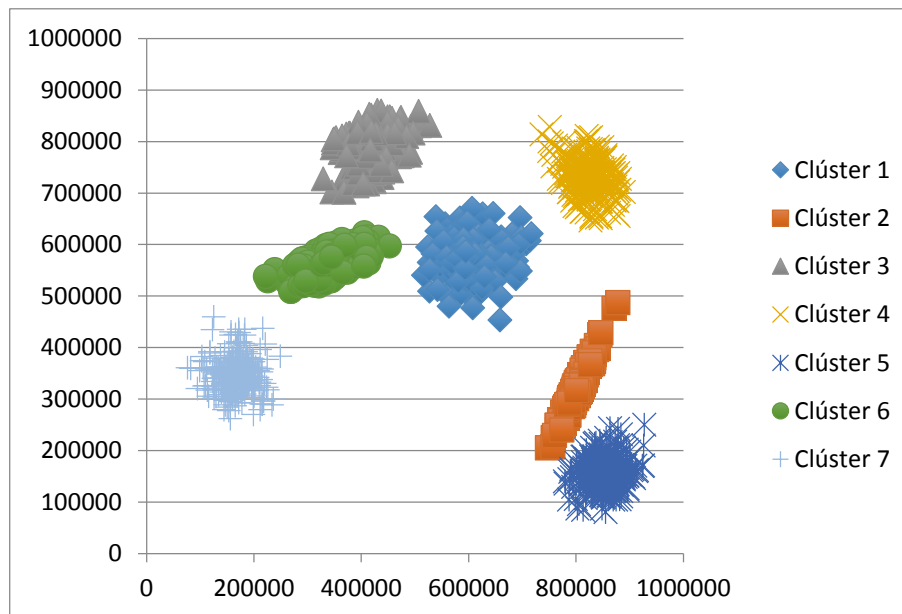


Figura 93. Conglomerados óptimos para el problema S1 truncado a 7 grupos

Para el problema S2 actuamos de forma similar. En esta ocasión sus grupos son menos homogéneos que en S1 y de nuevo lo truncamos para quedarnos con 3, 5 y 7 conglomerados. Tal como se aprecia en la Figura 95 y en la Figura 96 para 5 y 7 grupos, respectivamente, el solape es evidente entre algunos pares de grupos y es aquí donde las métricas tienen más problemas para discriminar entre unos y otros.

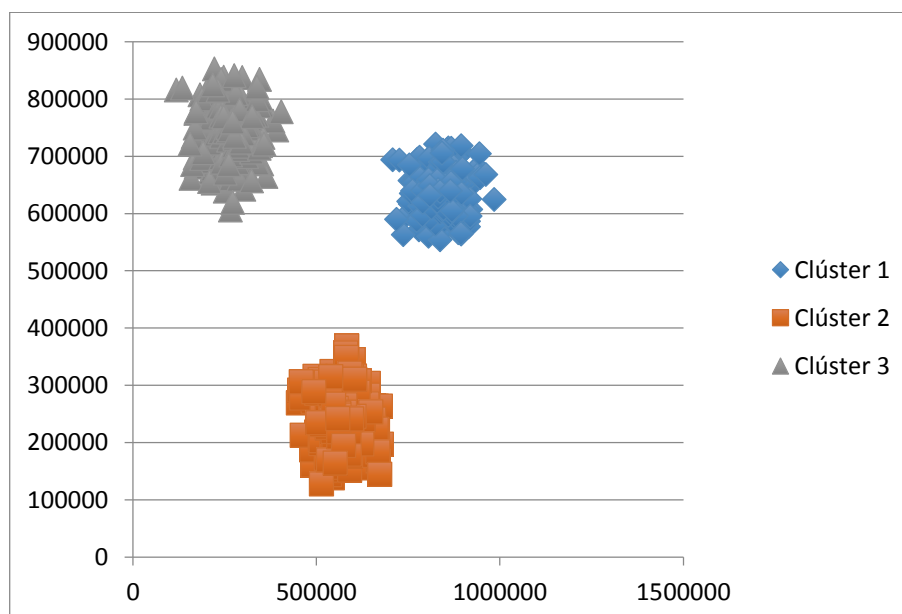


Figura 94. Conglomerados óptimos para el problema S2 truncado a 3 grupos

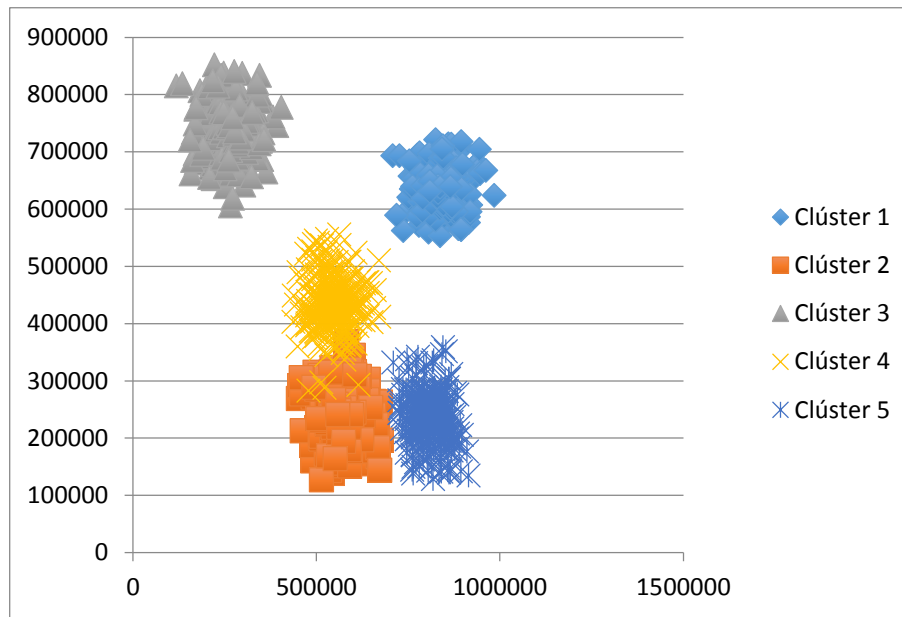


Figura 95. Conglomerados óptimos para el problema S2 truncado a 5 grupos

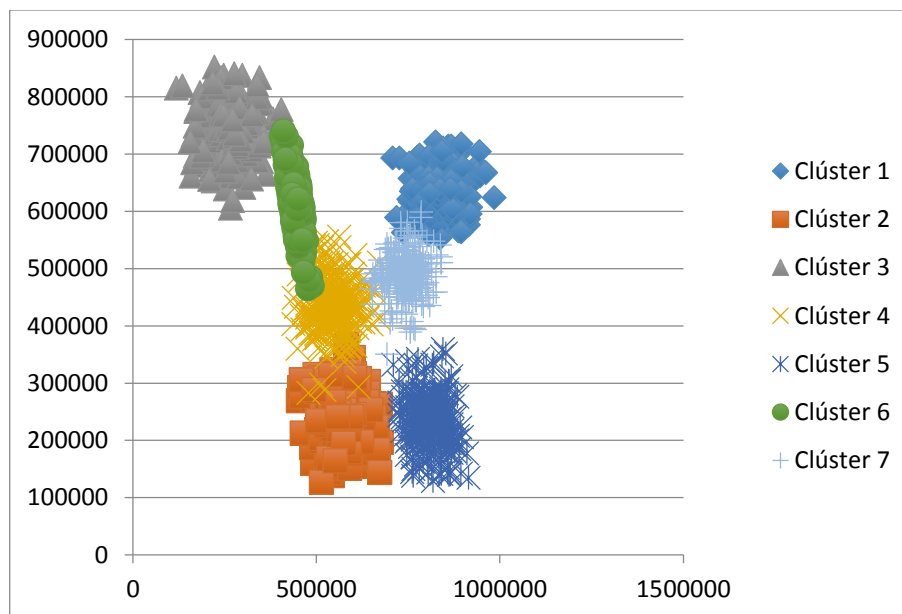


Figura 96. Conglomerados óptimos para el problema S2 truncado a 7 grupos

Y finalmente, el problema *S3* para 3, 5 y 7 grupos. En esta ocasión, no nos debería de extrañar que alguna métrica tenga grandes dificultades para detectar grupos distintos puesto que el nivel de solape es enorme. Es más, puesto que *DECCS* determina la mejor clasificación de manera automática para entre 2 y K_{max} clústeres, alguna métrica nos indicará que la mejor partición para estos problemas se obtiene con tan solo 2 grupos, en lugar de los 3, 5 ó 7. Por ello, en aquellos casos en los que el usuario necesite un número concreto de grupos, deberá usar otras técnicas o algoritmos para “forzar” esas divisiones poco naturales. Planteamos como trabajo futuro una adaptación de *DECCS* para este problema donde el número de grupos se conoce *a priori*.

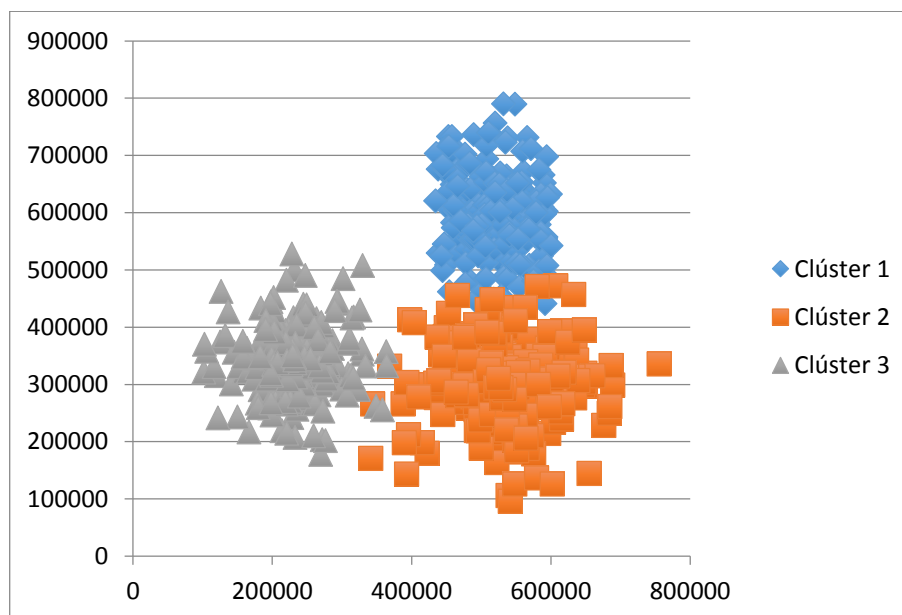


Figura 97. Conglomerados óptimos para el problema *S3* truncado a 3 grupos

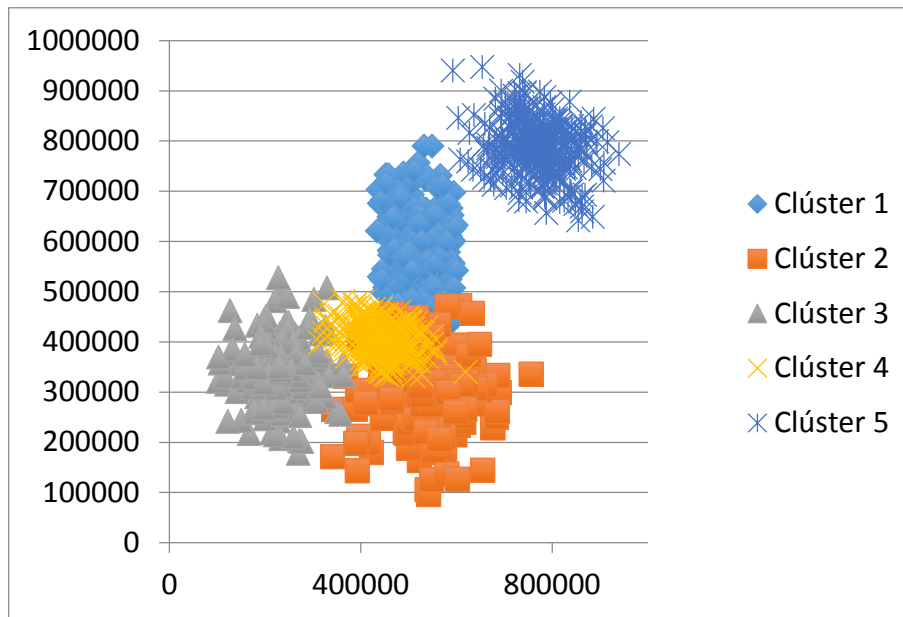


Figura 98. Conglomerados óptimos para el problema S3 truncado a 5 grupos

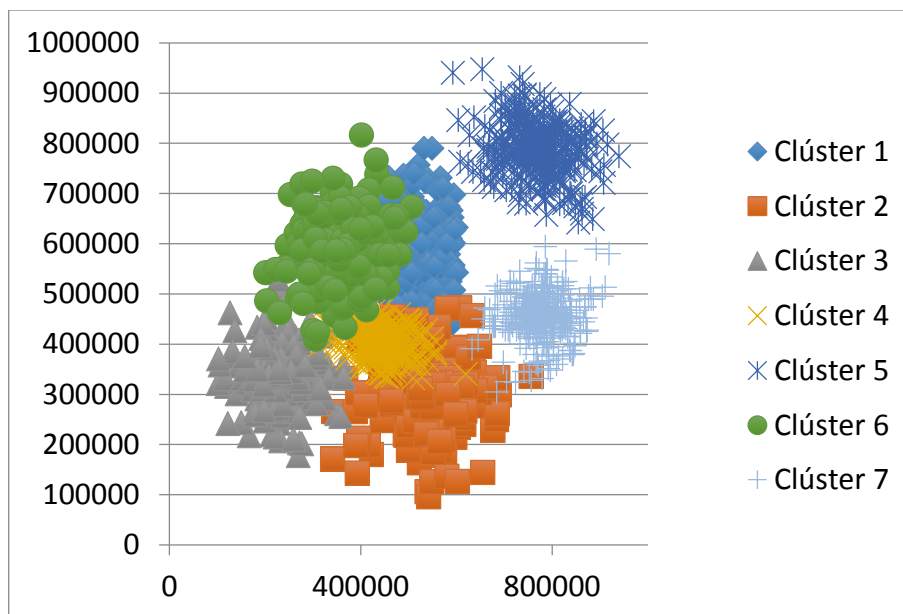


Figura 99. Conglomerados óptimos para el problema S3 truncado a 7 grupos

A continuación, resolvemos con nuestro algoritmo los nueve problemas para cada una de las cuatro métricas, obteniendo los siguientes resultados:

Resultados obtenidos para el problema S1

Los resultados obtenidos aparecen en la Tabla 49. En ella se aprecia como la métrica de *Calinsky-Harabasz* (CH) obtiene los mejores resultados pues obtiene los óptimos en los tres casos analizados. No obstante, destacamos en verde los resultados de la métrica CS que, sin ser los óptimos, presenta buenas propiedades y una sensibilidad que puede ser deseable, según el caso, a la hora de detectar grupos.

| Conjunto | Óptimos según la medida | | | | CS | | DN | | CH | | DB | |
|----------|-------------------------|--------|--------|--------|------|--------|------|--------|------|--------|------|--------|
| | CS | DN | CH | DB | # Cl | Valor | # Cl | Valor | # Cl | Valor | # Cl | Valor |
| S1_3 | 0,5505 | 0,6741 | 0,0003 | 0,5465 | 4 | 0,4933 | 3 | 0,6741 | 3 | 0,0003 | 3 | 0,5465 |
| S1_5 | 0,6512 | 1,1235 | 0,0003 | 0,8254 | 6 | 0,5373 | 2 | 0,8787 | 5 | 0,0003 | 3 | 0,6202 |
| S1_7 | 0,6363 | 1,1235 | 0,0003 | 0,8851 | 4 | 0,4955 | 4 | 0,9160 | 7 | 0,0003 | 3 | 0,6144 |

Tabla 49. Resultados obtenidos para el conjunto S1

Si analizamos con algo más de detalle los resultados de S1 apreciamos como para S1 con 3 grupos DN, CH y DB obtienen el óptimo. La métrica CS encuentra una agrupación “mejor” para cuatro grupos que para los tres “óptimos” (0,4933 frente a 0,5505), como también se observa en la Figura 100. ¿A qué se puede deber? Pues no es más que, según la medida de calidad usada CS, añadir un cuarto grupo con tan solo dos puntos mejora el valor de la métrica en comparación al valor de los tres grupos supuestamente óptimos. A esto nos referimos cuando afirmamos que la métrica CS es más sensible y genera un mayor número de grupos, frente a otras métricas que favorece a los grupos grandes. Se observa en la Figura 100 como la partición óptima se obtiene con cuatro grupos en lugar de los tres que eran los esperados.

Algo parecido ocurre con cinco grupos, donde CS indica que un sexto grupo para ese problema es mejor que los cinco grupos originales.

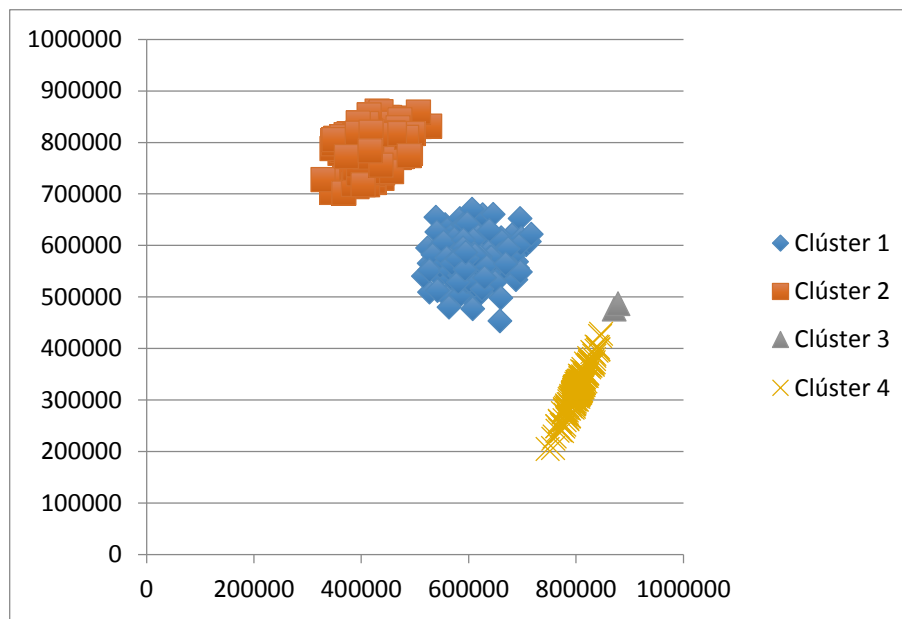


Figura 100. Solución obtenida por DECCS para S1 truncado para 3 grupos con la métrica CS

Resultados obtenidos para el problema S2

Para el problema S2 con 3, 5 y 7 grupos, éstos comienzan a solaparse levemente. Los resultados obtenidos se observan en la Tabla 50. En ella se aprecia como la métrica de *Calinsky-Harabasz* (CH) obtiene dos resultados óptimos pero una agrupación muy pobre para S2 de 7 grupos. En cambio, destacamos en verde los resultados de la métrica CS, que obtiene el óptimo para 3 grupos y mejores resultados que sus competidores para 5 y 7 grupos respectivamente con respecto al número de clústeres (#Cl).

| Conjunto | Óptimos según la medida | | | | CS | | DN | | CH | | DB | |
|----------|-------------------------|--------|--------|--------|------|--------|------|--------|------|--------|------|--------|
| | CS | DN | CH | DB | # Cl | Valor | # Cl | Valor | # Cl | Valor | # Cl | Valor |
| S2_3 | 0,3550 | 0,4009 | 0,0003 | 0,4577 | 3 | 0,3550 | 3 | 0,4009 | 3 | 0,0003 | 3 | 0,4577 |
| S2_5 | 0,6571 | 1,0113 | 0,0004 | 1,0084 | 4 | 0,5102 | 2 | 0,9830 | 5 | 0,0004 | 2 | 0,6279 |
| S2_7 | 0,8943 | 1,1285 | 0,0005 | 1,3773 | 5 | 0,5866 | 2 | 1,1421 | 2 | 0,0004 | 2 | 0,6278 |

Tabla 50. Resultados obtenidos para el conjunto S2

Resultados obtenidos para el problema S3

Finalmente, para el problema S3, con 3, 5 y 7 grupos que se solapan claramente, obtenemos unos resultados similares a los anteriores. De nuevo, aunque la métrica de *Calinsky-Harabasz* obtiene un resultado óptimo para 3 grupos, los resultados son muy pobres para 5 y 7 grupos. En cambio, destacamos en verde los resultados de la métrica CS que, sin obtener los óptimos, es más sensible a la hora de discriminar los datos y dividirlos en grupos más compactos para 3 y 5 grupos. Para S3 con 7 grupos las cuatro métricas arrojan valores malos. Esto era de esperar pues los 7 grupos se solapan enormemente y existe demasiado ruido en los datos como para poder discriminarlos y segmentarlos en los siete grupos supuestamente ideales.

Los resultados aparecen en la Tabla 51.

| | Óptimos según la medida | | | | CS | | DN | | CH | | DB | |
|----------|-------------------------|--------|--------|--------|------|----------|----|--------|----|--------|------|----------|
| Conjunto | CS | DN | CH | DB | # Cl | Conjunto | CS | DN | CH | DB | # Cl | Conjunto |
| S3_3 | 0,8470 | 0,9362 | 0,0007 | 1,1511 | 4 | 0,4925 | 3 | 0,8581 | 3 | 0,0007 | 2 | 0,6780 |
| S3_5 | 1,2093 | 1,9691 | 0,0007 | 1,7125 | 6 | 0,6190 | 2 | 0,9396 | 2 | 0,0005 | 2 | 0,6933 |
| S3_7 | 1,2086 | 1,9691 | 0,0007 | 1,9232 | 2 | 0,7095 | 2 | 1,2653 | 2 | 0,0004 | 2 | 0,7095 |

Tabla 51. Resultados obtenidos para el conjunto S2

En resumen, aunque son dos las métricas que destacan frente a las restantes, proponemos usar la métrica CS para validar el algoritmo pues, sin ser la que más óptimos ha obtenido, es la métrica más fina o sensible a la hora de discriminar y clasificar los datos. No obstante, y puesto que los tiempos de computación son pequeños, para problemas reales siempre es posible resolverlos con varias métricas y comparar los resultados obtenidos para elegir la mejor de ellas.

3.3 Simetría en la codificación

A continuación, mostraremos la validez de la codificación propuesta para DECCS. Aunque el desempeño de ACDE con su codificación original funcionó correctamente y se podría haber utilizado para DECCS, mostraremos que el esquema de reordenamiento para DECCS no solo mejora los resultados obtenidos sino que además acelera la obtención de los óptimos sin estancarse en óptimos locales. Para dicha validación se ha optado por usar ACDE y así no atribuir las mejoras obtenidas a otras aportaciones incorporadas a DECCS.

En la Tabla 52 comparamos ACDE sin reordenamiento con los resultados obtenidos por ACDE con reordenamiento y con la codificación propuesta. La estructura de la tabla es la siguiente: en la columna “Indicador”, mostramos la mejor (Min.), la peor (Max.), la ejecución media (Media) y la desviación estándar (DE) de las 40 resoluciones realizadas. De cada una de estas ejecuciones incluimos el valor de la función objetivo (*Func. Obj. (CS)*), el número de clústeres (# cl), las distancias intra e inter-clúster usada en *CS* (*Dist. Intra, Dist. Inter*), el número de evaluaciones realizadas (*Eval*), el tiempo empleado en cada ejecución en segundos (Tiempo (s)) y el p-valor obtenido al realizar un contraste de igualdad de medias de *Mann-Whitney*.

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | #cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|---|-----------------------------|-----------|-----------------|----------|-------------|------------|-------|----------------|-------------------|
| IRIS Variables = 4 Grupos = 3 Vectores = 150 | ACDE | Min. | 0,595638 | 3 | 5,3467 | 8,97643 | 20040 | 1,94961 | 1 |
| | | Media | 0,595638 | 3 | 5,3467 | 8,97643 | 20040 | 2,00237 | |
| | | Max. | 0,595638 | 3 | 5,3467 | 8,97643 | 20040 | 2,05232 | |
| | | DE | 1,97E-08 | 0 | 8,43E-08 | 3,15E-07 | 0 | 0,0202582 | |
| | ACDE con nueva codificación | Min. | 0,595638 | 3 | 5,3467 | 8,97643 | 20040 | 1,93628 | |
| | | Media | 0,595638 | 3 | 5,3467 | 8,97643 | 20040 | 2,0479 | |
| | | Max. | 0,595638 | 3 | 5,3467 | 8,97643 | 20040 | 2,14654 | |
| | | DE | 2,11E-08 | 0 | 0 | 1,69E-07 | 0 | 0,0429873 | |
| GLASS Variables = 9 Grupos = 6 Vectores = 214 | ACDE | Min. | 0,44445 | 3 | 9,3875 | 21,1216 | 45090 | 11,2256 | 1 |
| | | Media | 0,44445 | 3 | 9,3875 | 21,1216 | 45090 | 11,4939 | |
| | | Max. | 0,44445 | 3 | 9,3875 | 21,1216 | 45090 | 11,7804 | |
| | | DE | 0 | 0 | 1,19E-07 | 8,60E-07 | 0 | 0,118571 | |
| | ACDE con nueva codificación | Min. | 0,44445 | 3 | 9,3875 | 21,1216 | 45090 | 11,331 | |
| | | Media | 0,44445 | 3 | 9,3875 | 21,1216 | 45090 | 11,5582 | |
| | | Max. | 0,44445 | 3 | 9,3875 | 21,1216 | 45090 | 12,1185 | |
| | | DE | 9,13E-09 | 0 | 0 | 5,84E-07 | 0 | 0,15257 | |
| WINE Variables = 13 Grupos = 3 Vectores = 178 | ACDE | Min. | 0,60214 | 2 | 1055,55 | 1752,99 | 26130 | 9,79586 | 1 |
| | | Media | 0,60214 | 2 | 1055,55 | 1752,99 | 26130 | 10,0097 | |
| | | Max. | 0,60214 | 2 | 1055,55 | 1752,99 | 26130 | 10,2352 | |
| | | DE | 0 | 0 | 0 | 5,29E-05 | 0 | 0,103674 | |
| | ACDE con nueva codificación | Min. | 0,60214 | 2 | 1055,55 | 1752,99 | 26130 | 3,8283 | |
| | | Media | 0,60214 | 2 | 1055,55 | 1752,99 | 26130 | 3,89927 | |
| | | Max. | 0,60214 | 2 | 1055,55 | 1752,99 | 26130 | 4,01091 | |
| | | DE | 0 | 0 | 0 | 0 | 0 | 0,031216 | |
| BREAST Variables = 9 Grupos = 2 Vectores = 699 | ACDE | Min. | 0,982633 | 2 | 23,7906 | 24,2111 | 45090 | 80,6096 | 0,0000024 |
| | | Media | 1,14468 | 2,02 | 33,8215 | 29,5654 | 45090 | 81,2779 | |
| | | Max. | 1,31053 | 3 | 40,9526 | 32,6843 | 45090 | 82,9528 | |
| | | DE | 0,066559 | 0,14 | 2,48645 | 1,75509 | 0 | 0,347596 | |
| | ACDE con nueva codificación | Min. | 0,863863 | 2 | 27,8856 | 26,2437 | 45090 | 81,0575 | |
| | | Media | 1,04189 | 2 | 32,6067 | 31,3909 | 45090 | 81,7993 | |
| | | Max. | 1,23491 | 2 | 36,0138 | 33,8967 | 45090 | 82,8174 | |
| | | DE | 0,092643 | 0 | 2,37538 | 1,83346 | 0 | 0,355884 | |
| THYROID Variables = 5 Grupos = 3 Vectores = 215 | ACDE | Min. | 0,414411 | 4 | 80,3406 | 191,396 | 25050 | 5,38193 | 0,00354049 |
| | | Media | 0,428557 | 4 | 82,9308 | 193,516 | 25050 | 5,45512 | |
| | | Max. | 0,43584 | 4 | 84,7457 | 194,576 | 25050 | 5,55468 | |
| | | DE | 0,00577 | 0 | 1,02052 | 0,845343 | 0 | 0,0393564 | |
| | ACDE con nueva codificación | Min. | 0,414411 | 4 | 80,3406 | 191,396 | 25050 | 5,54789 | |
| | | Media | 0,425581 | 4 | 82,5107 | 193,878 | 25050 | 5,64335 | |
| | | Max. | 0,428065 | 4 | 83,0035 | 193,958 | 25050 | 5,77745 | |
| | | DE | 0,005144 | 0 | 1,01334 | 0,398848 | 0 | 0,0525426 | |

Tabla 52. Resultados comparativos entre ACDE sin reordenamiento y ACDE con reordenamiento y con la codificación propuesta en problemas de tamaño medio

Comenzamos con los problemas de tamaño medio. De los cinco problemas resueltos, se obtiene que no existen diferencias significativas entre las medias para *IRIS*, *GLASS* y *WINE* ($p\text{-valor}=1$) mientras que para los problemas *BREAST* y *THYROID* sí se aprecian diferencias a favor del esquema con reordenamiento pues los $p\text{-valores}$ son inferiores a 0,005 en ambos casos.

No obstante, aunque los valores promedios no presentan diferencias significativas en sus valores finales para *IRIS*, *GLASS* y *WINE*, las velocidades de convergencia de ambos algoritmos son muy distintas, como se puede apreciar en las figuras comprendidas entre la Figura 101 y la Figura 105.

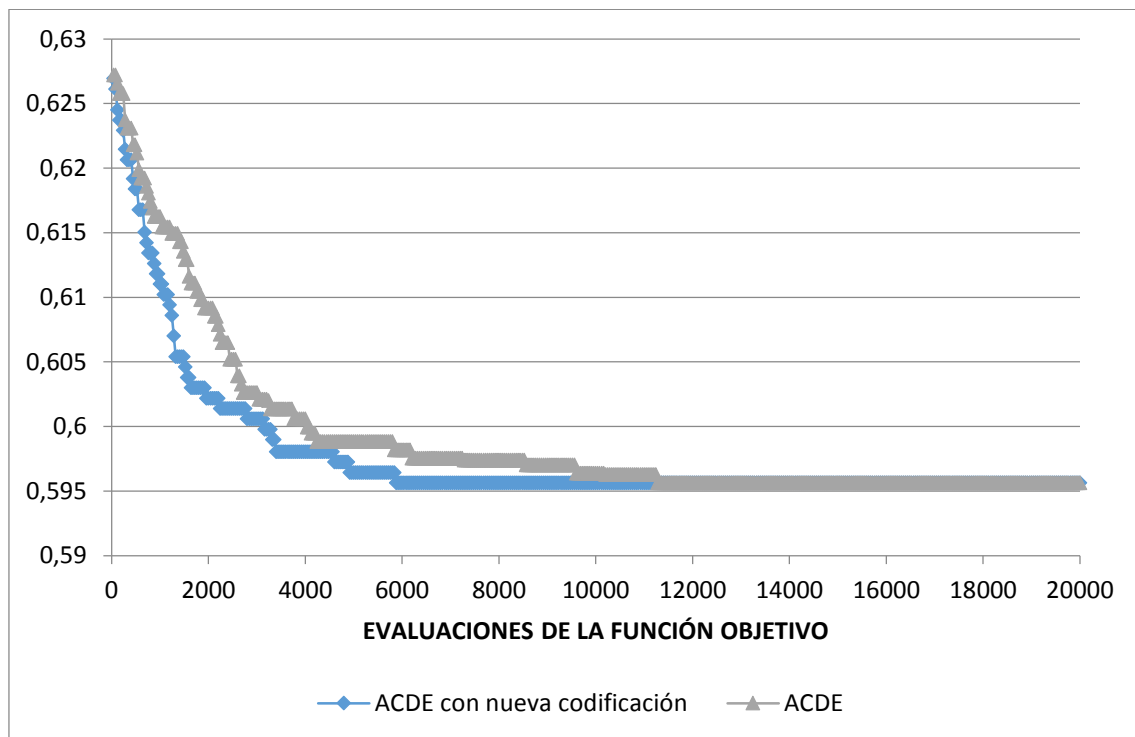


Figura 101. Velocidades de convergencia media para *IRIS*

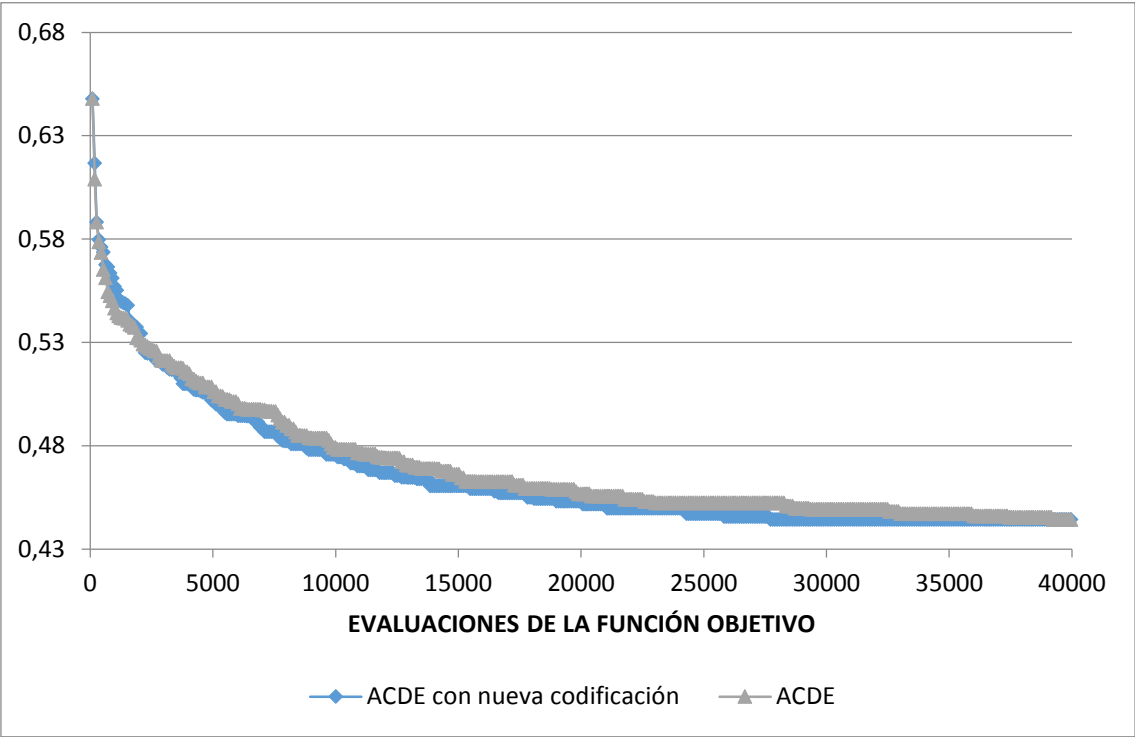


Figura 102. Velocidades de convergencia media para GLASS

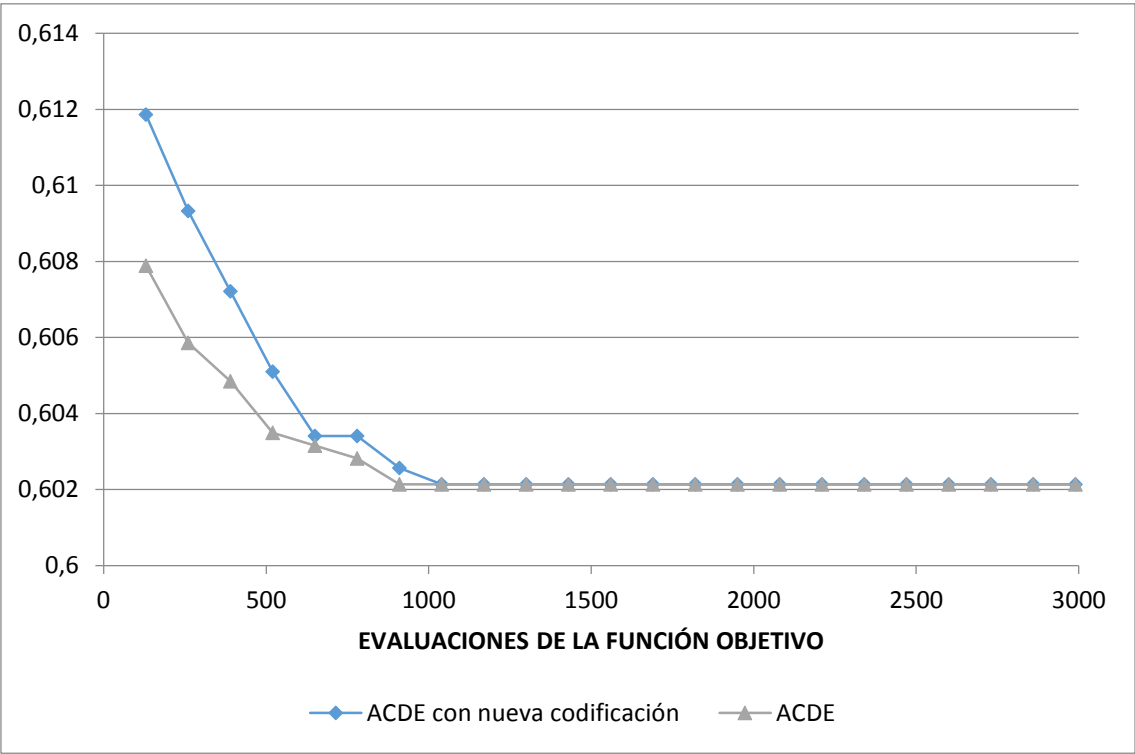


Figura 103. Velocidades de convergencia media para WINE

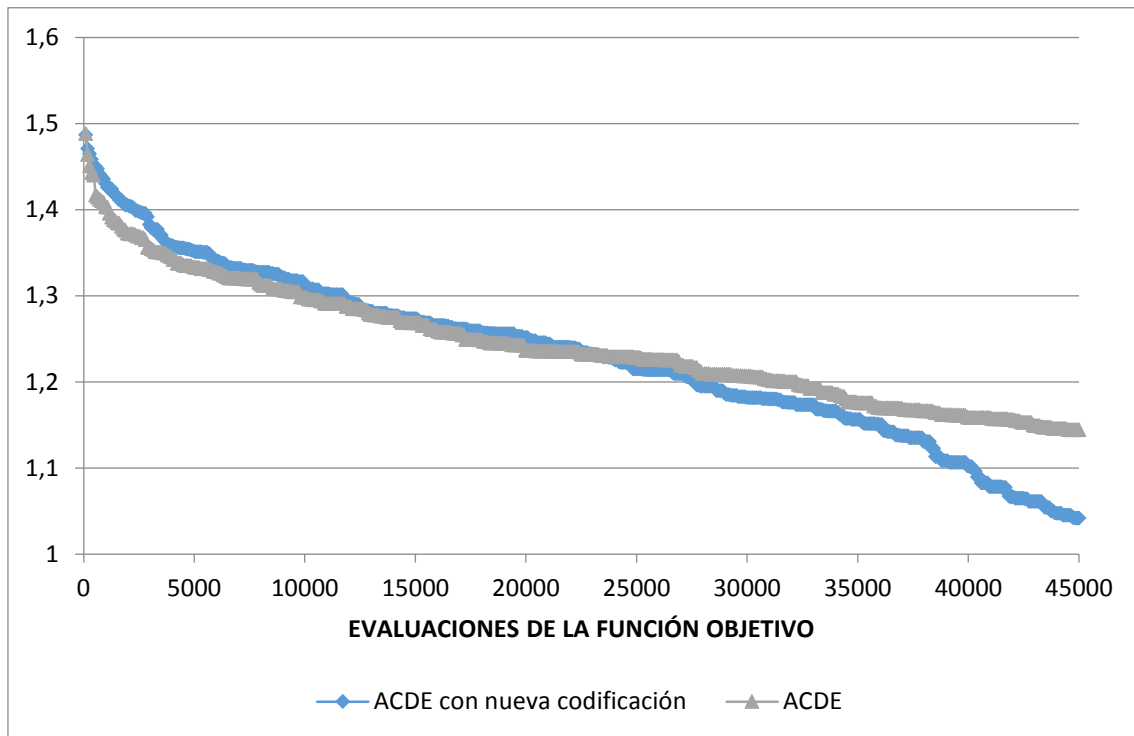


Figura 104. Velocidades de convergencia media para BREAST

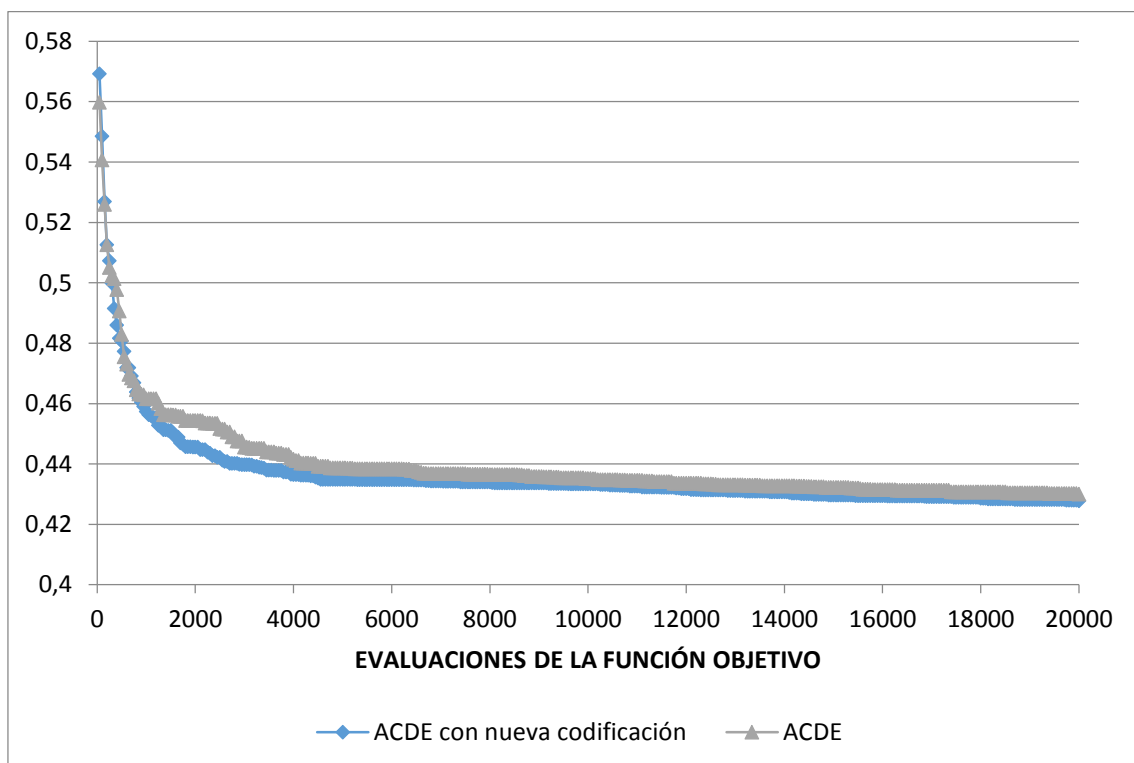


Figura 105. Velocidades de convergencia media para THYROID

Pasemos a analizar los resultados para los problemas de mayor complejidad. Ahora, el p-valor obtenido a favor del esquema de reordenamiento es casi 0 en tres de los cuatro problemas (*LED*, *YEAST* y *SONAR*) y un p-valor igual a 1 para *WINE QUALITY*. De nuevo, encontramos diferencias significativas en los rendimientos de ACDE y ACDE con reordenamiento. Los resultados aparecen en la Tabla 53.

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|--|-----------------------------|-----------|-----------------|---------|-------------|------------|--------|------------|-----------|
| LED Var = 7 Grupos = 10 Vec = 500 | ACDE | Min. | 1,1310 | 13 | 17,1719 | 14,3404 | 35070 | 37,8015 | 0,0000206 |
| | | Media | 1,1949 | 15,8750 | 20,5601 | 17,2065 | 35070 | 38,2603 | |
| | | Max. | 1,2374 | 20 | 24,0225 | 20,0502 | 35070 | 39,0072 | |
| | | DE | 0,0273 | 1,646 | 1,8128 | 1,4661 | 0 | 0,2732 | |
| | ACDE con nueva codificación | Min. | 1,0424 | 13 | 17,2605 | 14,3296 | 35070 | 40,7113 | |
| | | Media | 1,1590 | 16,8500 | 21,1693 | 18,2938 | 35070 | 41,3611 | |
| | | Max. | 1,2388 | 20 | 24,6082 | 21,3210 | 35070 | 42,1503 | |
| | | DE | 0,0416 | 1,6820 | 1,6824 | 1,6388 | 0 | 0,3230 | |
| YEAST Var = 8 Grupos = 10 Vec = 1484 | ACDE | Min. | 0,4397 | 2 | 0,9636 | 1,8514 | 40080 | 335,8160 | 0,0000024 |
| | | Media | 0,4870 | 3,5750 | 1,4076 | 2,8976 | 40080 | 338,9860 | |
| | | Max. | 0,5205 | 5 | 1,9810 | 3,8774 | 40080 | 340,9960 | |
| | | DE | 0,0225 | 0,6280 | 0,1883 | 0,4124 | 0 | 1,2454 | |
| | ACDE con nueva codificación | Min. | 0,3897 | 3 | 1,0287 | 2,3282 | 40080 | 336,4220 | |
| | | Media | 0,4606 | 3,8000 | 1,4162 | 3,0807 | 40080 | 342,4780 | |
| | | Max. | 0,5015 | 5 | 1,7951 | 3,9314 | 40080 | 346,9490 | |
| | | DE | 0,0194 | 0,7140 | 0,2005 | 0,4650 | 0 | 2,4419 | |
| WINE QUALITY Var = 11 Grupos = 11 (6) Vec = 1599 | ACDE | Min. | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 185,3530 | 1 |
| | | Media | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 187,4980 | |
| | | Max. | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 190,9470 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0,0000 | 0 | 1,3026 | |
| | ACDE con nueva codificación | Min. | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 180,9930 | |
| | | Media | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 184,7780 | |
| | | Max. | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 186,7520 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0,0000 | 0 | 1,3885 | |
| SONAR Var = 60 Grupos = 2 Vec = 208 | ACDE | Min. | 0,7419 | 2 | 3,1898 | 3,9587 | 600600 | 220,0170 | 0,0000958 |
| | | Media | 0,7981 | 2,4250 | 3,7501 | 4,7071 | 600600 | 221,4510 | |
| | | Max. | 0,8281 | 3 | 4,6768 | 5,8572 | 600600 | 223,0750 | |
| | | DE | 0,0161 | 0,4940 | 0,6331 | 0,8389 | 0 | 0,6027 | |
| | ACDE con nueva codificación | Min. | 0,7019 | 2 | 3,1898 | 3,9832 | 600600 | 221,1900 | |
| | | Media | 0,7803 | 2,7000 | 4,0241 | 5,1695 | 600600 | 222,6500 | |
| | | Max. | 0,8010 | 3 | 4,6033 | 5,8485 | 600600 | 224,3720 | |
| | | DE | 0,0240 | 0,4580 | 0,5728 | 0,7874 | 0 | 0,7087 | |

Tabla 53. Resultados comparativos entre ACDE sin reordenamiento y ACDE con reordenamiento y con la codificación propuesta en problemas de gran tamaño

Si analizamos las velocidades de convergencia media para los cuatro problemas, aunque en especial nos interesa el problema *WINE QUALITY*, apreciamos en las figuras comprendidas entre la Figura 106 y Figura 108 los beneficios del uso del esquema de reordenamiento frente a no usarlo. En el conjunto de datos *WINE QUALITY*, ambos algoritmos encuentran el óptimo en las primeras iteraciones por lo que no se aprecia beneficio alguno.

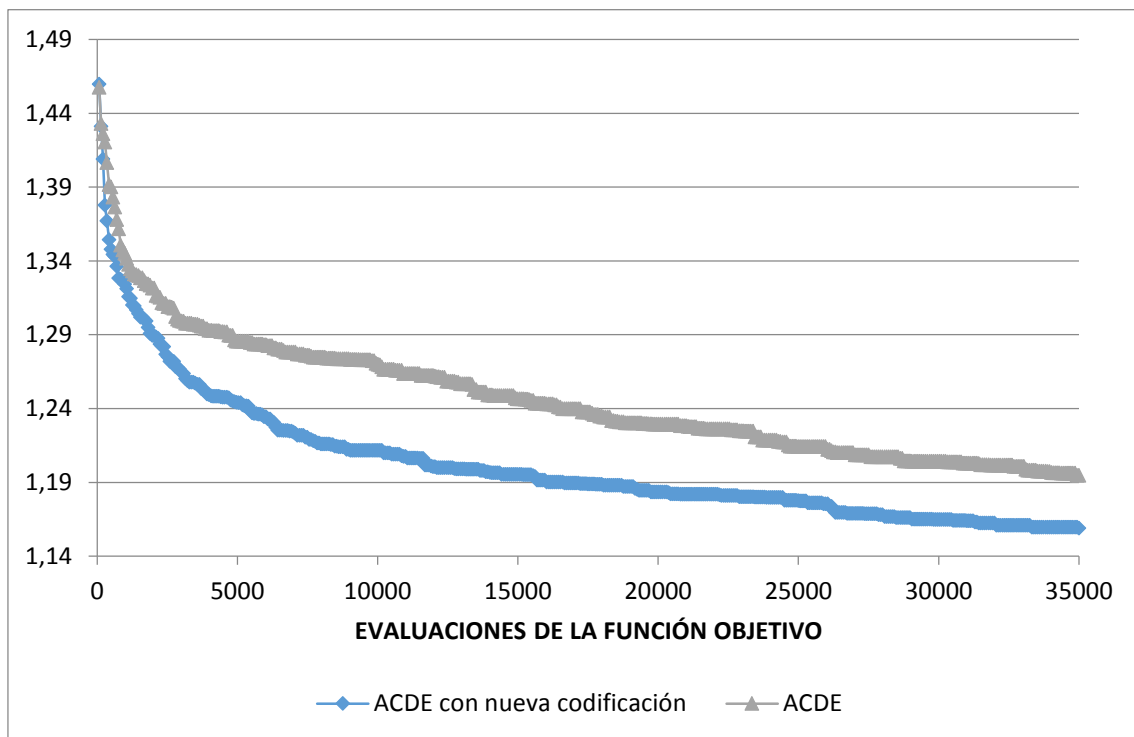


Figura 106. Velocidades de convergencia media para LED

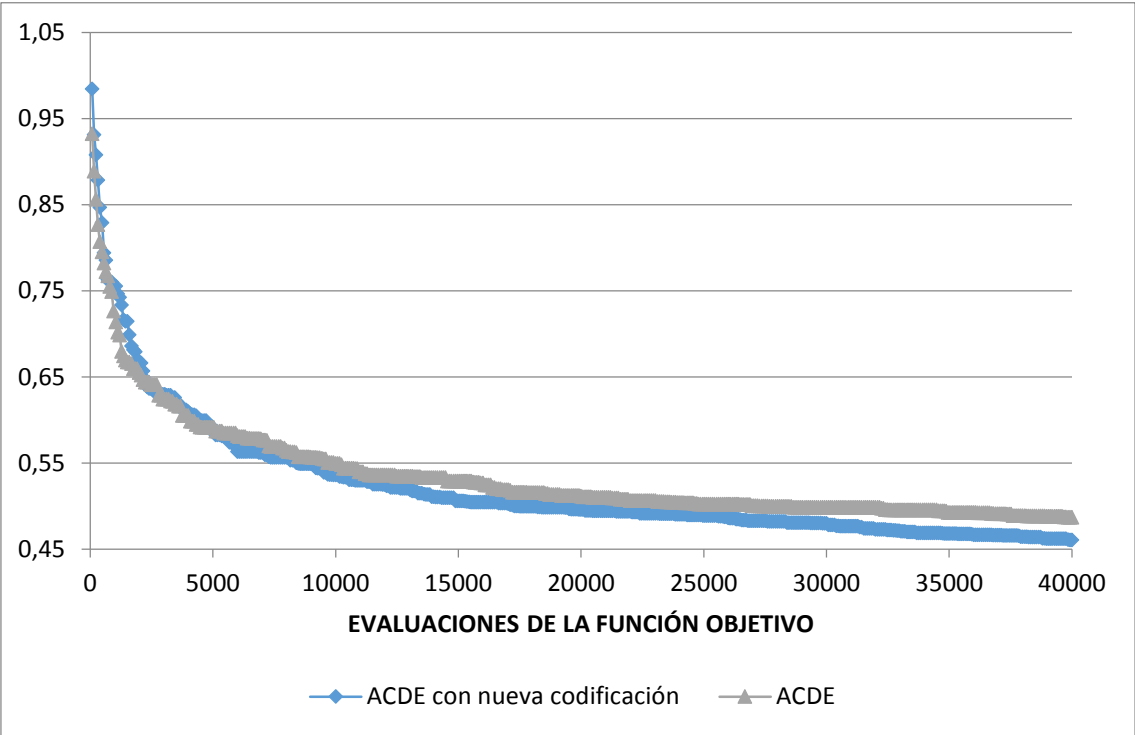


Figura 107. Velocidades de convergencia media para YEAST

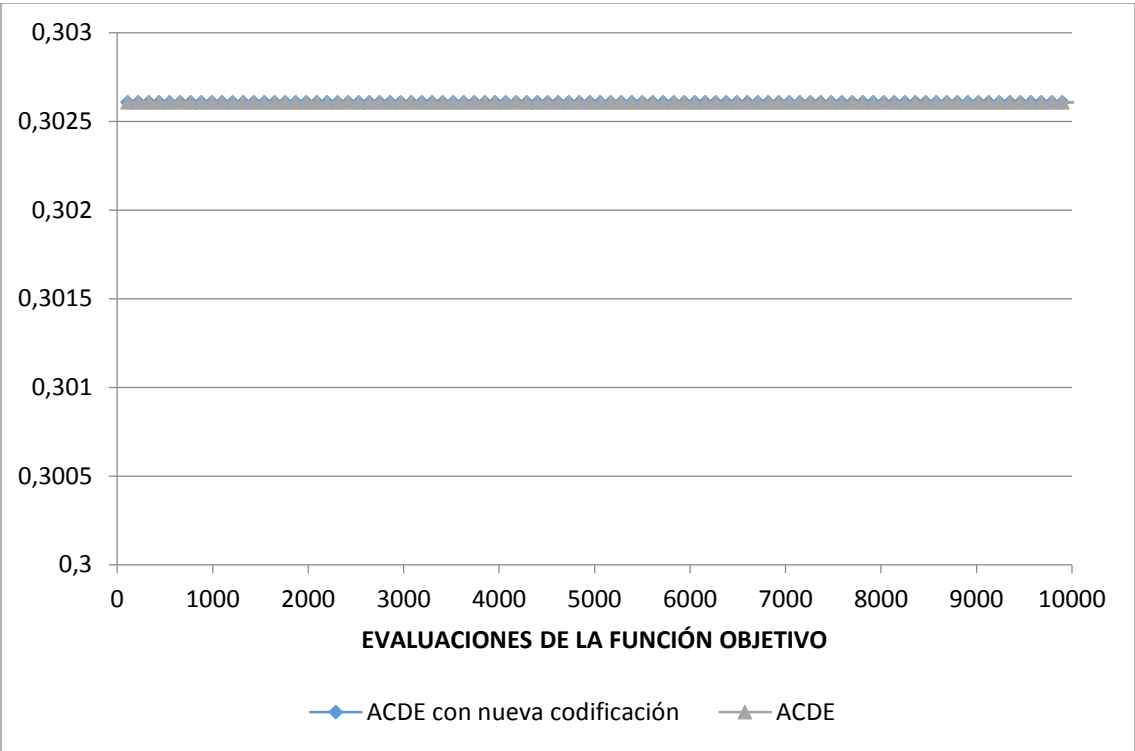


Figura 108. Velocidades de convergencia media para WINE QUALITY

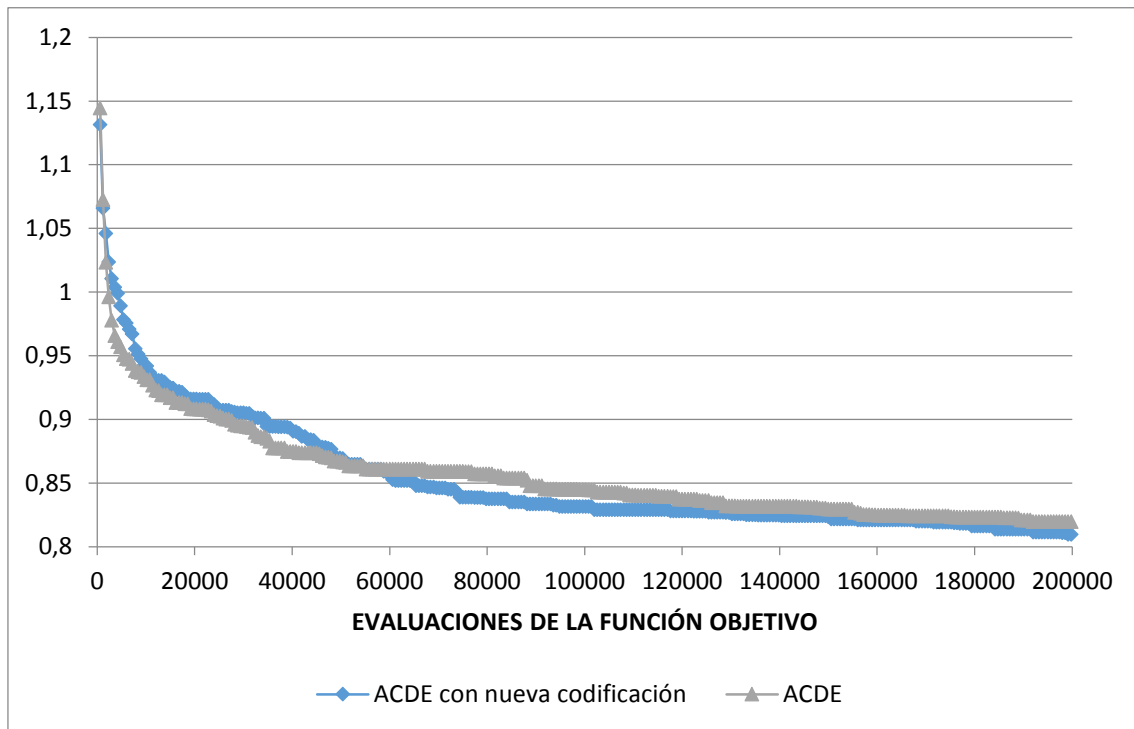


Figura 109. Velocidades de convergencia media para SONAR

Conclusión

El esquema de reordenamiento es claramente beneficioso en especial para problemas de gran dimensión puesto que acelera la convergencia sin perder diversidad en la población.

3.4 Comparativa de métodos de muestreo para la población inicial

En este tercer bloque de experimentos, analizaremos la capacidad exploradora de tres técnicas de muestreo para problemas de segmentación con la codificación propuesta para DECCS.

Como ya se ha resaltado previamente, la construcción de una población inicial de calidad, a la vez que diversa, es fundamental para que un algoritmo evolucione adecuadamente y no converja prematuramente. Los tamaños de las poblaciones suelen ser pequeños (entre 20 y 100 soluciones) y, por tanto, es deseable que al generarlo no se dejen sin explorar grandes zonas del espacio. Así, al generar dichos puntos o soluciones se debe evitar generar puntos en zonas cercanas a otros puntos previamente generados, así como cubrir en la medida de lo posible, los rangos de valores de las variables.

A su vez, el tamaño de la población inicial es también clave a la hora de medir el rendimiento de un algoritmo evolutivo. Obviamente, mientras mayor es el tamaño de la población más zonas del espacio se pueden explorar, más diversidad se tendrá en la población y mejores soluciones iniciales se podrán obtener. No obstante, a medida que el tamaño de la población crece, el coste computacional de una evolución diferencial de cualquier algoritmo evolutivo se incrementa exponencialmente. Por otro lado, tamaños de población reducidos se traducen en algoritmos más rápidos pero también a pérdida de diversidad y a convergencia prematura en los algoritmos. Este dilema y otros relacionados pueden ser estudiados con mayor profundidad en (207), (208), (209) y (210). En especial, destacamos el trabajo de Segura y col. (207), donde se proponen los dos mecanismos que mejoran la capacidad exploradora de un algoritmo evolutivo para tamaños de población reducidos e incorporados en DECCS y cuyo uso para problemas de segmentación proponemos en esta disertación. Así, la combinación de tamaños pequeños de población (también conocidos como microalgoritmos) con técnicas como las incorporadas en DECCS da lugar a un algoritmo rápido a la vez que robusto y eficiente.

Las tres técnicas seleccionadas en esta sección, muy populares en computación evolutiva, son las siguientes:

- Hipercubos latinos (en inglés, *Latin-hypercubes*).
- Vectores opuestos (en inglés, *Opposite-based arrays*).
- Aleatoria (en inglés, *Naive method*).

La técnica de hipercubos latinos ya fue detallada en la sección 2.1.4. Las dos nuevas técnicas se detallan a continuación. Antes de ello, y para facilitar la explicación de los métodos, supondremos que el problema requiere de $NVar$ variables, x_i , que varían en los intervalos $x_i \in [cota_{inf_i}, cota_{sup_i}]$.

Vectores opuestos

El aprendizaje basado en vectores opuestos fue presentado por Tizhoosh en 2005 (213). La clave principal, detrás de esta novedosa idea, es la consideración simultánea de una estimación, o muestra en nuestro caso, y su correspondiente estimación opuesta, a fin de lograr una mejor aproximación a la solución. Posteriormente, la consideración de individuos opuestos en el espacio de variables para la generación de la población inicial fue empleado exitosamente con Evolución Diferencial en multitud de trabajos. Alguno de ellos puede verse en (214), (215), (216), (217) y (218).

El concepto de *opuesto* aparece en numerosas áreas de estudio pero para nuestro propósito la idea se puede exponer de manera simple para una sola variable y generalizarla para un espacio multidimensional.

Definición: sea el intervalo $[a, b]$ donde se selecciona aleatoriamente un valor x siguiendo una distribución uniforme tal que $x \in [a, b]$. El valor opuesto de x en el intervalo se denota por \check{x} y se calcula como $\check{x} = a + b - x$.

En la Figura 110 se ofrece la interpretación geométrica para una dimensión y en la Figura 111 para dos dimensiones.

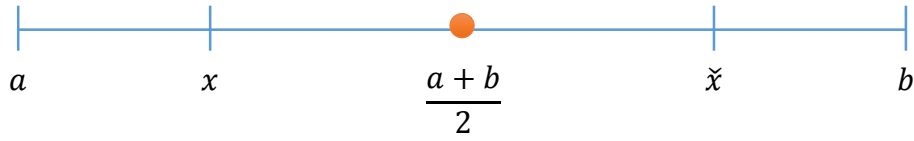


Figura 110. Vectores opuestos en la recta real

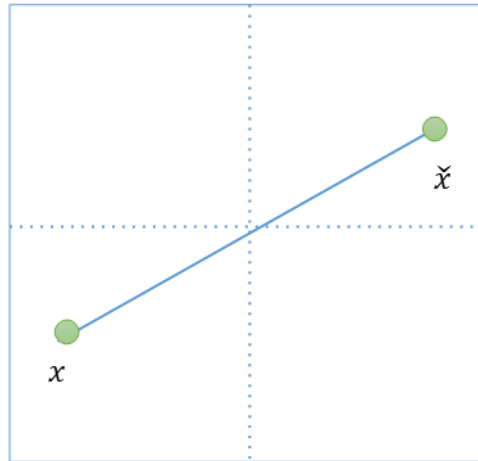


Figura 111. Vectores opuestos en el plano

Aleatorio

Es el método más simple de los tres y consiste en seleccionar puntos aleatoriamente en el espacio de variables usando una distribución uniforme multidimensional sin más restricciones. Así, cada vector $x = (x_1, x_2, \dots, x_{NVar})$ se genera mediante el uso de distribuciones uniformes en los intervalos $x_i \in [cota_{inf_i}, cota_{sup_i}]$. Aunque su formulación es sencilla y aparentemente válida, estos métodos tienen dificultades para generar puntos en los extremos del espacio y tienden a concentrar sus puntos en la zona central.

Pasamos a continuación a determinar no solo el mejor de los semillados para nuestros problemas de segmentación, sino a la vez el tamaño de población para éste, que equilibre la capacidad de exploración con el coste computacional. Debemos tener en cuenta que DECCS incorpora dos esquemas que lo ayudan a explorar nuevas zonas del espacio a pesar de manejar poblaciones pequeñas. Por este motivo, nuestra intención

es mostrar también que el método que finalmente se seleccione funciona adecuadamente con tamaños pequeños.

Con este fin vamos a generar $TamPop = 20, 30, 40$ y 100 soluciones en una selección representativa de cuatro problemas. El criterio para seleccionar estos cuatro problemas ha sido el tamaño del espacio de soluciones y fueron:

- IRIS. La codificación necesaria para este problema requiere de **30 variables** pues posee 4 variables y 3 grupos.
- GLASS. La codificación necesaria para este problema requiere de **120 variables** pues posee 9 variables y 6 grupos.
- S1. La codificación necesaria para este problema requiere de **45 variables** pues posee 2 variables y 15 grupos.
- YEAST. La codificación necesaria para este problema requiere de **190 variables** pues posee 8 variables y 10 grupos.

Estas soluciones son evaluadas y comparadas respecto a la métrica seleccionada. Recordemos que se está usando la métrica CS aunque los resultados obtenidos en esta investigación no dependen de la métrica.

Puesto que existe una alta componente estocástica en las tres técnicas expuestas, se generarán 51 poblaciones iniciales para cada problema y tamaño y se procederá a comparar con los correspondientes contrastes estadísticos. Así, se han realizado un total de: $4 \text{ tamaños} \times 4 \text{ problemas} \times 51 \text{ ejecuciones} = 1.804$ comparaciones.

A continuación, se detallan los resultados obtenidos de las comparaciones realizadas (véase el Anexo 6.2 para consultar todas las ejecuciones realizadas).

Problema IRIS

En la Figura 112 se observa la evolución de la mejor, la peor y la media de la mejor solución encontrada en 51 muestreos del problema para 20, 30, 40 y 100 soluciones para un muestreo aleatorio. En la Figura 113 y en la Figura 114 se observa la evolución para el caso de hipercubos latinos y vectores opuestos respectivamente.

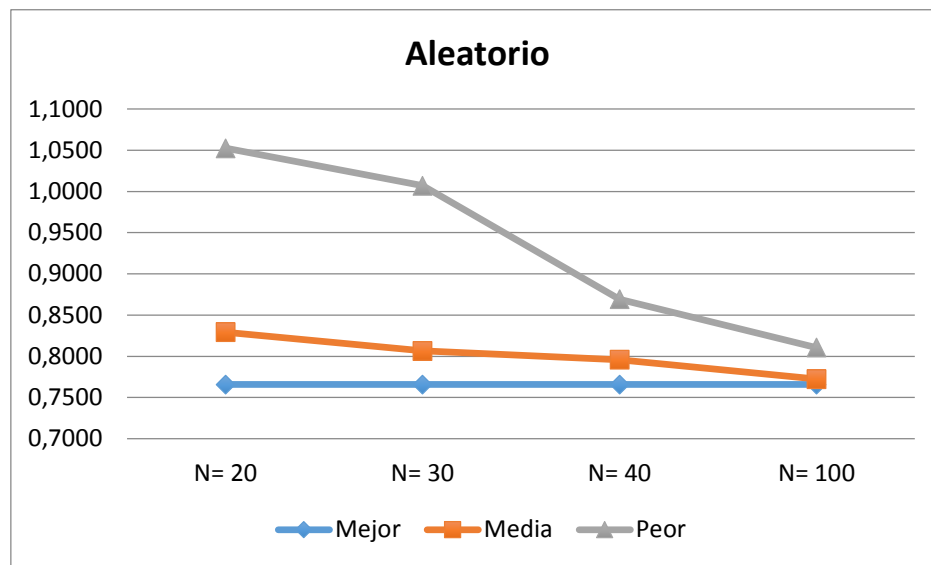


Figura 112. Calidad media, mejor y peor valor obtenido para 51 muestreos aleatorios para el problema IRIS

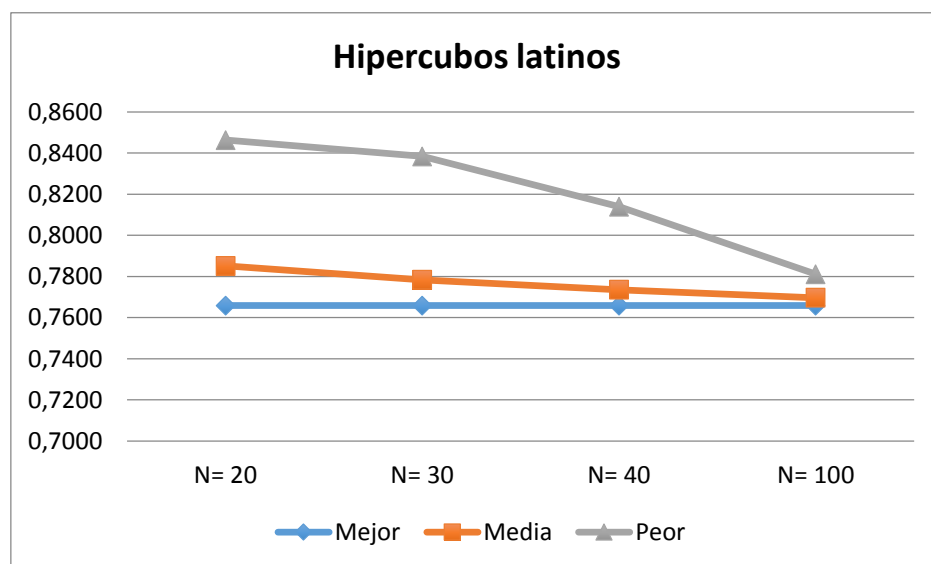


Figura 113. Calidad media, mejor y peor valor obtenido para 51 muestreos usando Hipercubos latinos para el problema IRIS

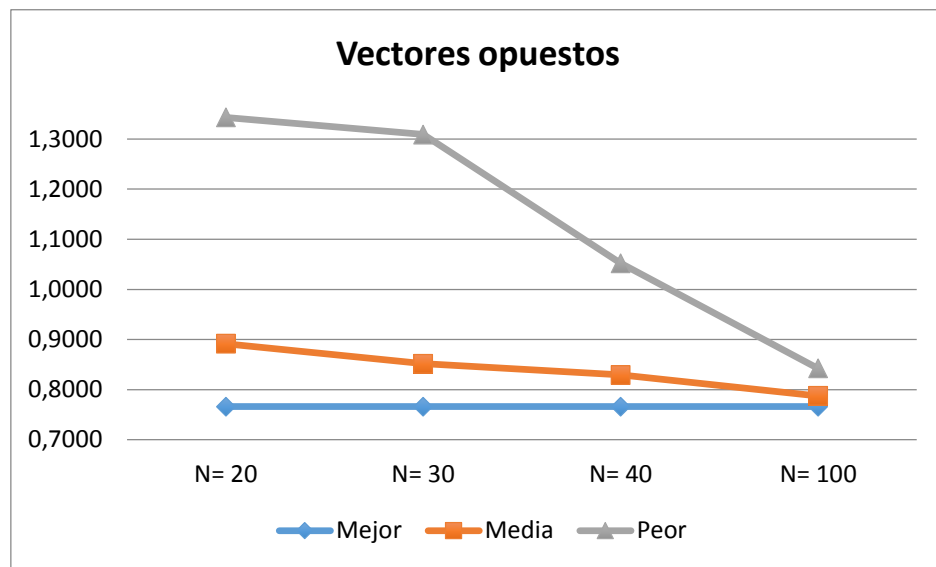


Figura 114. Calidad media, mejor y peor valor obtenido para 51 muestreos usando vectores opuestos para el problema IRIS

Como era de esperar, se aprecian tendencias decrecientes en todos los casos aunque los valores medio (en color naranja) son muy estables entre los cuatro tamaños analizados. La representación de los valores promedio obtenidos por las tres técnicas para los distintos tamaños de población de soluciones se muestra en la Figura 115.

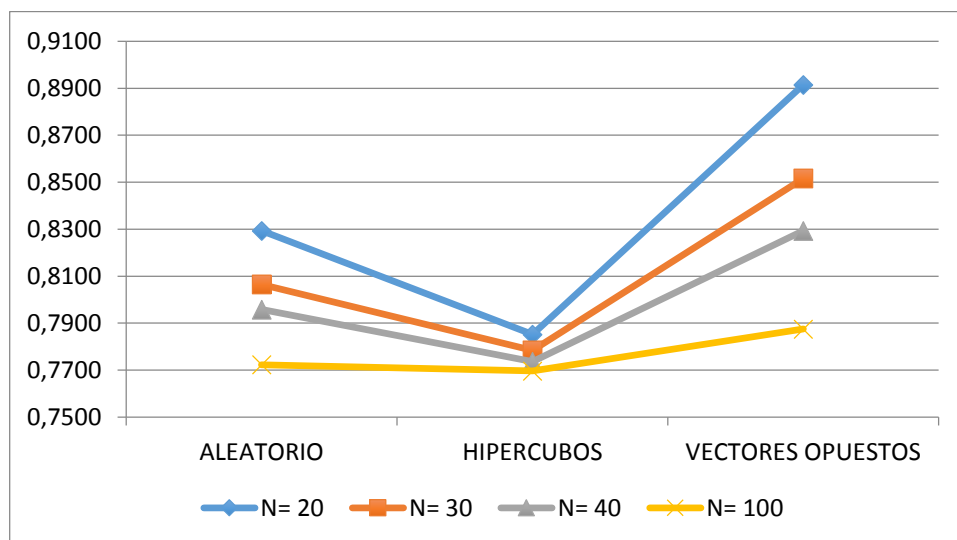


Figura 115. Calidad media para 51 muestreos usando los tres métodos propuestos para el problema IRIS con tamaños de población 20, 30, 40 y 100

Claramente los valores promedio son siempre mejores para los hipercubos latinos frente a la generación aleatoria de soluciones y la basada en vectores opuestos. Si realizamos

un contraste de *Kruskal-Wallis* para los cuatro tamaños del muestreo basado en hipercubos latinos se obtiene un p-valor = 0,000000538 (ver Tabla 54). Por tanto, existen diferencias significativas entre las medias obtenidas para 20, 30, 40 y 100 soluciones. No obstante las diferencias no son tan grandes como con los otros dos métodos por lo que el uso de tamaños de 20 soluciones no parece una mala elección.

| | |
|--------------------|-------------|
| <i>adjusted H:</i> | 31,943 |
| <i>d.f.:</i> | 3 |
| <i>P value:</i> | 0,000000538 |

Tabla 54. Resultados del contraste Kruskal-Wallis para los cuatro tamaños del muestreo basado en hipercubos latinos para el problema IRIS

Conclusión

Los resultados obtenidos son contundentes a favor de un muestreo basado en hipercubos latinos frente al aleatorio y al basado en vectores opuestos.

Problema GLASS

Los resultados obtenidos son muy similares al problema anterior. Por un lado se obtienen valores medios muy estables para distintos valores y por otro una superioridad clara para los hipercubos latinos frente a los otros dos métodos.

En la Figura 116 se muestra la evolución de la mejor, la peor y la media de la mejor solución encontrada en 51 muestreos del problema para 20, 30, 40 y 100 soluciones para un muestreo aleatorio. En la Figura 117 y en la Figura 118 se muestra la evolución para el caso de hipercubos latinos y vectores opuestos respectivamente.

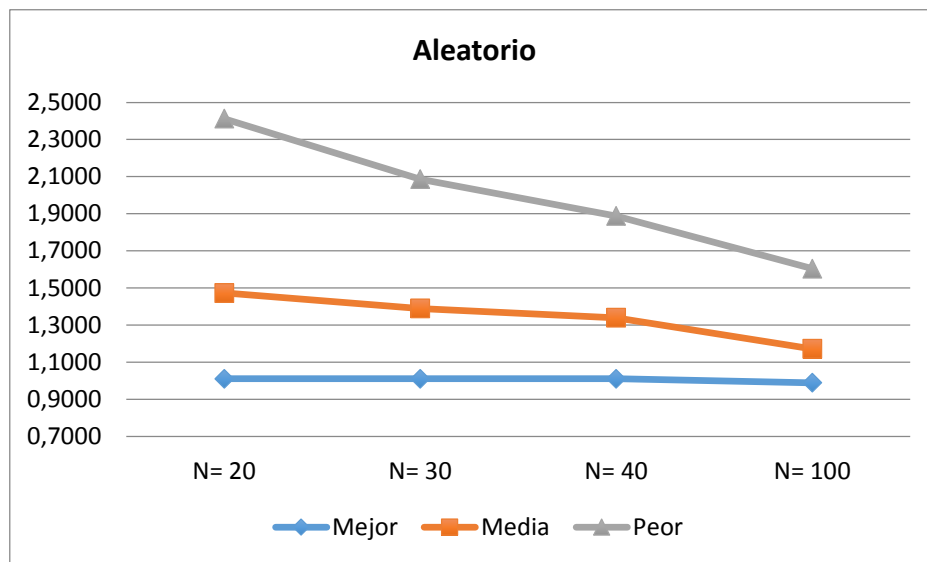


Figura 116. Calidad media, mejor y peor valor obtenido para 51 muestreos aleatorios para el problema GLASS

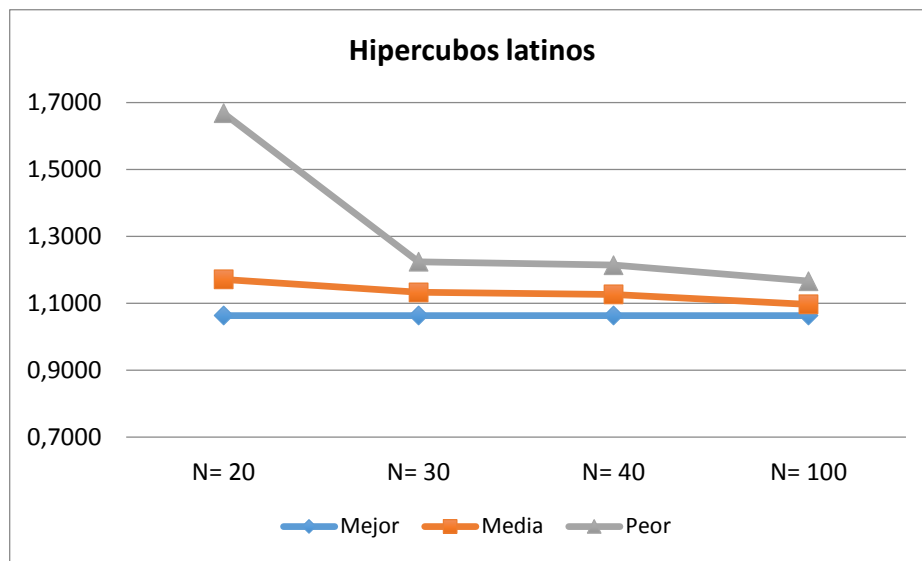


Figura 117. Calidad media, mejor y peor valor obtenido para 51 muestreos usando hipercubos latinos para el problema GLASS

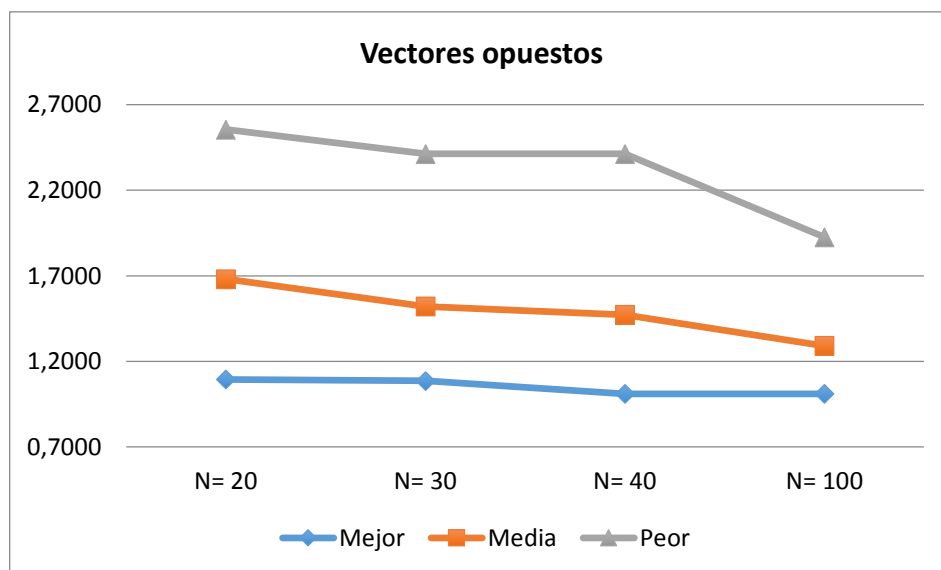


Figura 118. Calidad media, mejor y peor valor obtenido para 51 muestreos usando vectores opuestos para el problema GLASS

La representación de los valores promedio obtenidos por las tres técnicas para los distintos tamaños de población de soluciones se aprecia en la Figura 119.

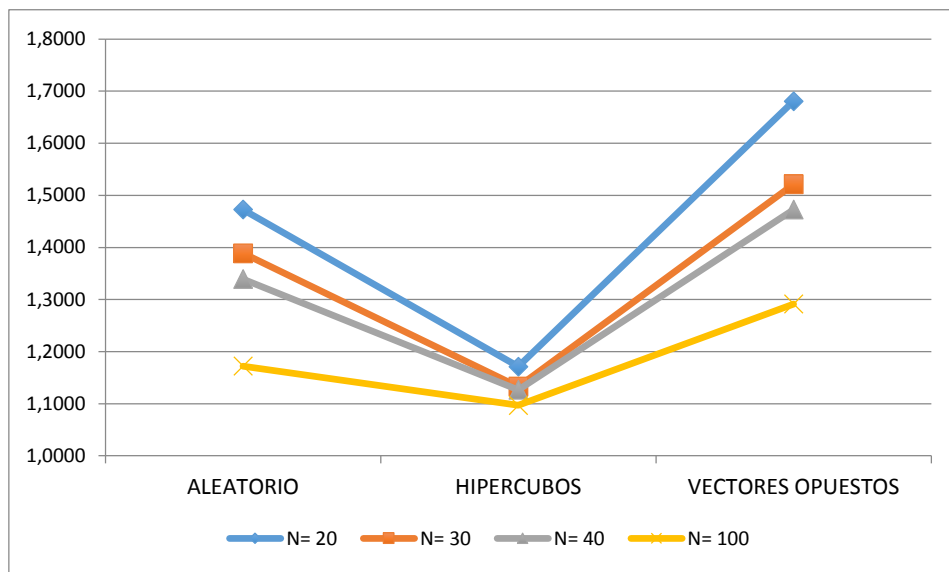


Figura 119. Calidad media para 51 muestreos usando los tres métodos propuestos para el problema GLASS con tamaños de población 20, 30, 40 y 100

De nuevo realizamos un contraste no paramétrico de *Kruskal-Wallis* para los cuatro tamaños usados para los hipercubos latinos con idea de buscar diferencias significativas entre las medias de los cuatro tamaños. Los resultados para el problema de *GLASS* fueron en esta ocasión más claros aunque, como se aprecia en la Figura 117, la calidad media para 20 soluciones está próxima a la de tamaños superiores.

| | |
|--------------------|-----------|
| <i>adjusted H:</i> | 118,219 |
| <i>d.f.:</i> | 3 |
| <i>P value:</i> | 1,866E-25 |

Tabla 55. Resultados del contraste *Kruskal-Wallis* para los cuatro tamaños del muestreo basado en hipercubos latinos para el problema *GLASS*

Conclusión

Los resultados obtenidos vuelven a estar a favor de un muestreo basado en hipercubos latinos frente al aleatorio y al basado en vectores opuestos.

Problema S1

Para S1 se obtuvieron resultados similares. Mostrando los hipercubos latinos una superioridad aún mayor frente a los otros dos métodos.

En la Figura 120 se aprecia la evolución de la mejor, la peor y la media de la mejor solución encontrada en 51 muestreos del problema para 20, 30, 40 y 100 soluciones para un muestreo aleatorio. En la Figura 121 y en la Figura 122 se muestra la evolución para el caso de hipercubos latinos y vectores opuestos respectivamente.

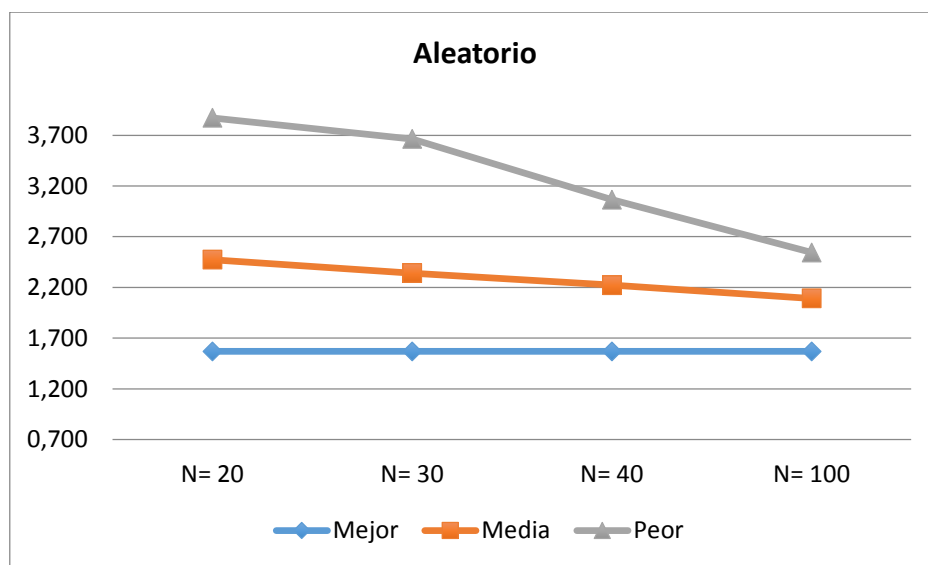


Figura 120. Calidad media, mejor y peor valor obtenido para 51 muestreos aleatorios para el problema S1

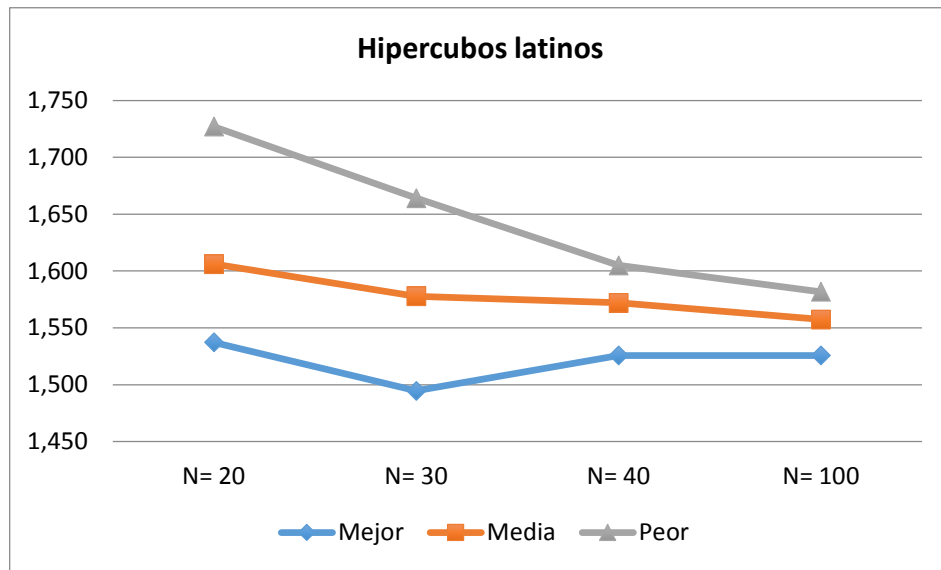


Figura 121. Calidad media, mejor y peor valor obtenido para 51 muestreos usando hipercubos latinos para el problema S1

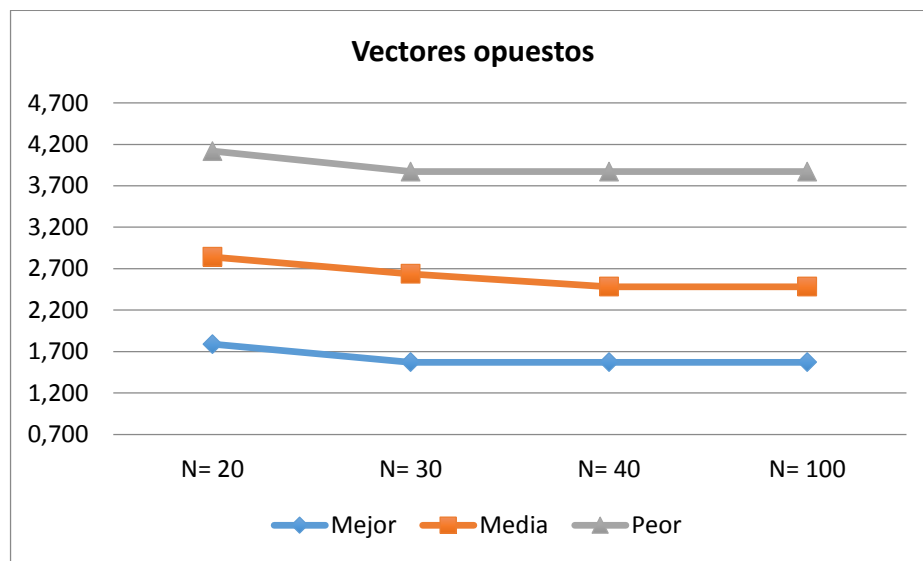


Figura 122. Calidad media, mejor y peor valor obtenido para 51 muestreos usando vectores opuestos para el problema S1

La representación de los valores promedio obtenidos por las tres técnicas para los distintos tamaños de población de soluciones se observa en la Figura 123.

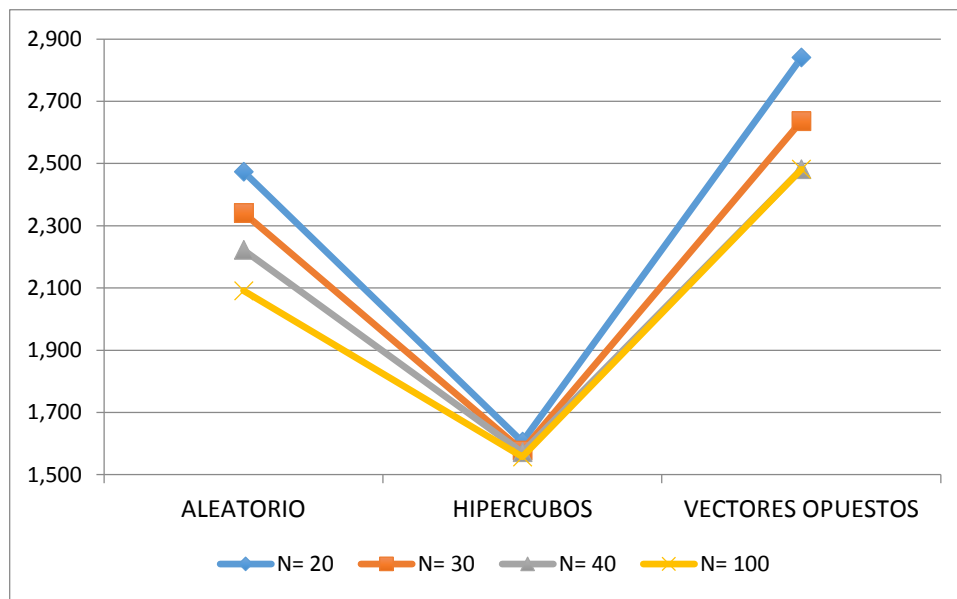


Figura 123. Calidad media para 51 muestreos usando los tres métodos propuestos para el problema S1 con tamaños de población 20, 30, 40 y 100

El contraste de *Kruskal-Wallis* en esta ocasión arroja un p-valor de nuevo muy cercano a 0, por lo que existen diferencias significativas entre los cuatro tamaños de muestra. De nuevo las diferencias entre 20 y el resto de tamaños es pequeña y dicha diferencia es asumible para DECCS puesto que ahorrará mucho tiempo de computación sin perder calidad.

| | |
|-------------|-----------|
| adjusted H: | 72,566 |
| d.f.: | 3 |
| P value: | 1,204E-15 |

Tabla 56. Resultados del contraste *Kruskal-Wallis* para los cuatro tamaños del muestreo basado en hipercubos latinos para el problema S1

Conclusión

Los resultados obtenidos vuelven, como en los casos anteriores, a estar a favor de un muestreo basado en hipercubos latinos frente al aleatorio y al basado en vectores opuestos. En esta ocasión, la diferencia es más acusada.

Problema YEAST

Por último, confirmamos los resultados mostrados hasta ahora con un problema de 190 variables de codificación.

En la Figura 124 se observa la evolución de la mejor, la peor y la media de la mejor solución encontrada en 51 muestreos del problema para 20, 30, 40 y 100 soluciones para un muestreo aleatorio. En la Figura 125 y en la Figura 126 se muestra la evolución para el caso de hipercubos latinos y vectores opuestos respectivamente.

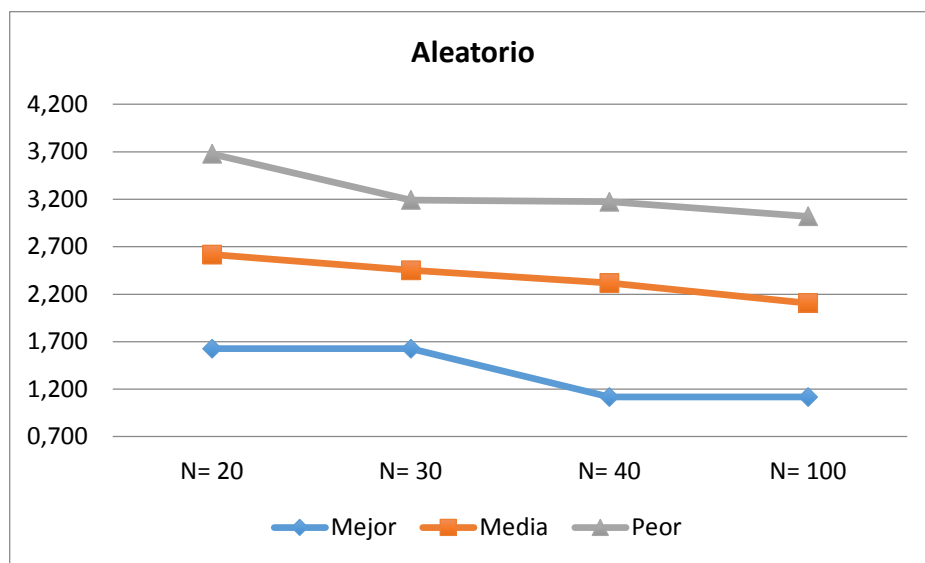


Figura 124. Calidad media, mejor y peor valor obtenido para 51 muestreos aleatorios para el problema YEAST

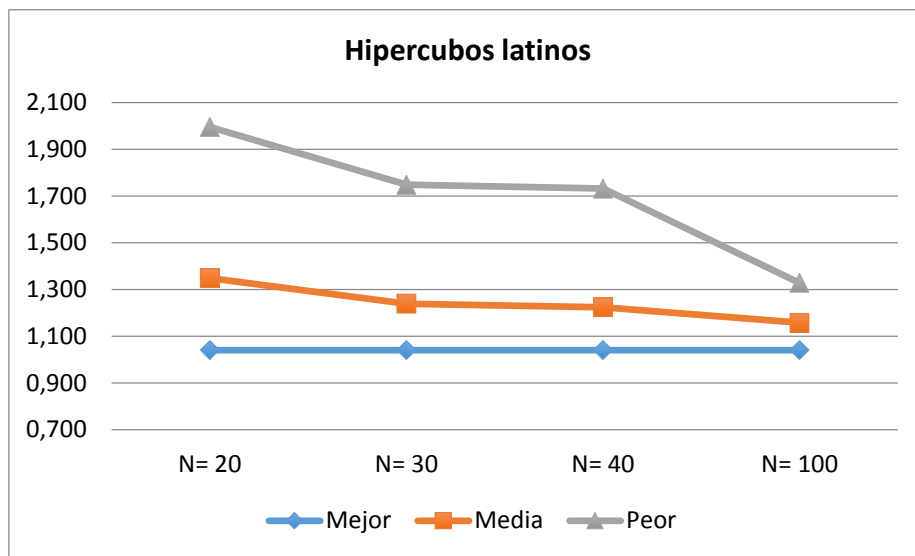


Figura 125. Calidad media, mejor y peor valor obtenido para 51 muestreos usando hipercubos latinos para el problema YEAST

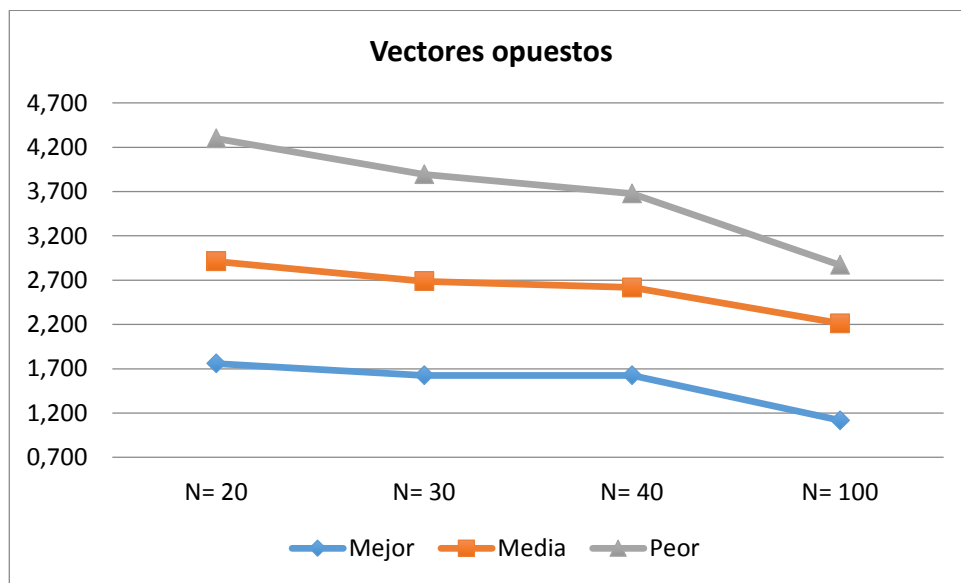


Figura 126. Calidad media, mejor y peor valor obtenido para 51 muestreos usando vectores opuestos para el problema YEAST

La representación de los valores promedio obtenidos por las tres técnicas para los distintos tamaños de población de soluciones se ofrece en la Figura 127.

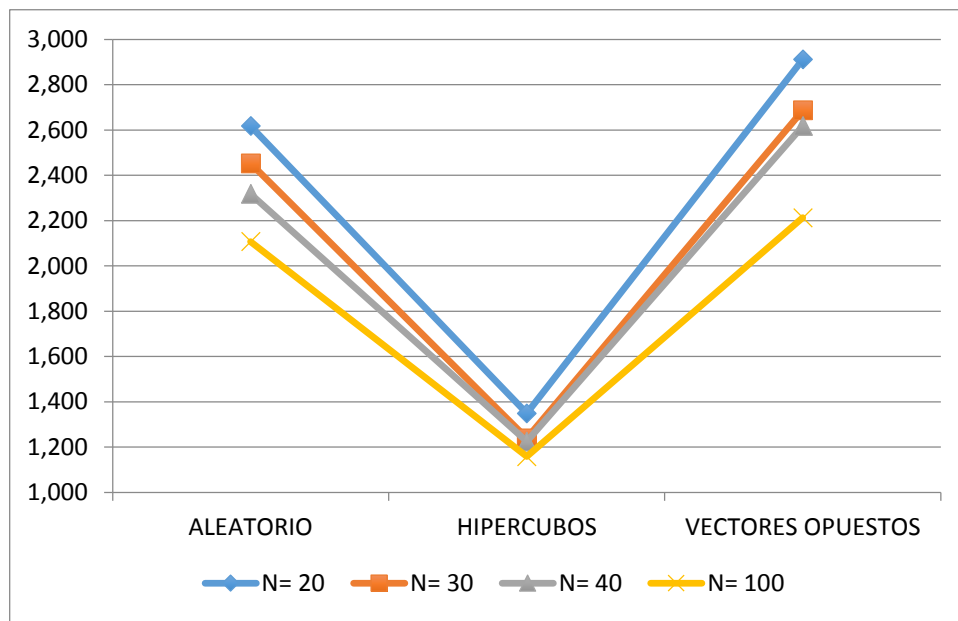


Figura 127. Calidad media para 51 muestreos usando los tres métodos propuestos para el problema YEAST con tamaños de población 20, 30, 40 y 100

En esta ocasión el p-valor es casi nulo por lo que se aprecian diferencias entre los cuatro tamaños.

| | |
|--------------------|----------|
| adjusted H: | 40,263 |
| d.f.: | 3 |
| P value: | 9,370E-9 |

Tabla 57. Resultados del contraste Kruskal-Wallis para los cuatro tamaños del muestreo basado en hipercubos latinos para el problema YEAST

Conclusión

Los resultados son claros a favor del método basado en hipercubos latinos. Además, el uso de un tamaño 20 para la población no hacer empeorar mucho el rendimiento medio por lo este será el valor que usaremos de ahora en adelante.

3.5 Comparativa de DECCS frente a otros algoritmos de segmentación

En este bloque final mostramos los resultados obtenidos por DECCS en las bases de datos de tamaño medio y grande, en un problema muy grande con cerca de 13.500 datos (problema *PEN*) y en un problema real obtenido de una oficina de farmacia.

Dichos resultados son comparados frente a los siguientes algoritmos:

- ACDE. Es su competidor natural pues la codificación usada por DECCS está inspirada en la propuesta de ACDE. Algoritmo formulado por Swagatam y col. en el año 2009 (134).
- *K-means*. Algoritmo que ya se expuso en la sección 1.4.7.1.1.
- Bietápico de *SPSS*. Expuesto en la sección 1.4.7.2.5.
- Jerárquico. Algoritmo expuesto con en la sección 1.4.7.2.

Los parámetros utilizados por ACDE y DECCS se detallan en la Tabla 58. Aunque los tamaños de población son claramente distintos, el número de iteraciones utilizadas por ambos algoritmos se ha ajustado para que los dos usen el mismo número de evaluaciones de la función objetivo. Cada problema se ha resuelto 40 veces y se acompañan con los correspondientes p-valores obtenidos tras aplicar un contraste no paramétrico de *Mann-Withney*.

| | ACDE | DECCS |
|-------------------------------------|--|---|
| Tamaño de población (TamPop) | 10*NVar | 20 |
| Población inicial | Aleatoria | Hipercubos latinos |
| Mutación y Cruce | Binario y exponencial | Binario y exponencial con esquema de continuación |
| F | Decreciente de 1 a 0.5 | |
| Esquema de continuación | No usa | Parámetros de los autores Segura y col. (207) |
| Kmax | 2 * número de grupos óptimos conocidos | |

Tabla 58. Parámetros de ejecución para los algoritmos ACDE y DECCS

3.5.1 Comparativa de resultados

Los resultados van en concordancia con otros previamente mostrados. En ellos veremos que los resultados se pueden dividir en dos grupos:

- Problemas donde no existen diferencias significativas entre las medias pero sí entre las velocidades de convergencia. Es el caso de *IRIS* y *GLASS* pues en *WINE* se muestra ACDE más rápido que DECCS aunque ambos acaban convergiendo al óptimo. Solo un problema de gran tamaño cae en esta categoría, *WINE QUALITY*. A pesar de su tamaño, no es complejo para ninguno de los dos algoritmos por lo que no requiere de técnicas más elaboradas para su resolución.
- Problemas donde sí existen diferencias significativas a favor de DECCS con p-valores casi nulos. Es el caso de *BREAST* y *THYROID* para problemas de tamaño medio y *LED*, *YEAST* y *SONAR* de gran tamaño. Destacamos aquí la superioridad de DECCS en tres de los cuatro problemas de gran tamaño con lo que se muestra un algoritmo solvente y eficaz para problemas de mucho mayor tamaño.

Las tablas que mostramos a continuación para los nueve problemas incluyen una comparativa de los resultados obtenidos por DECCS y ACDE y las soluciones obtenidas por *K-means*, bietápico y jerárquico (implementados en *SPSS*). Estos tres algoritmos se han ejecutado tanto para el valor óptimo del número de grupos o segmentos como para el valor entero más próximo obtenido por DECCS y ACDE. Como se verá, la calidad de sus soluciones según la métrica *CS* es peor en todos los casos.

Problema IRIS

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|---|------------|-----------|-----------------|--------|-------------|------------|-------|---------------|---------|
| IRIS Var = 4 Grupos = 3 Vec = 150 | DECCS | Min. | 0,5926 | 3 | 5,3467 | 8,9764 | 20020 | 3,6823 | 1 |
| | | Media | 0,5956 | 3,0250 | 5,3495 | 8,9823 | 20020 | 3,8870 | |
| | | Max. | 0,5956 | 4 | 5,4586 | 9,2109 | 20020 | 4,0893 | |
| | | DE | 0,0005 | 0,1561 | 0,0175 | 0,0366 | 0 | 0,0905 | |
| | ACDE | Min. | 0,5956 | 3 | 5,3467 | 8,9764 | 20040 | 1,9496 | |
| | | Media | 0,5956 | 3 | 5,3467 | 8,9764 | 20040 | 2,0024 | |
| | | Max. | 0,5956 | 3 | 5,3467 | 8,9764 | 20040 | 2,0523 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0 | 0 | 0,0203 | |
| | K-means | k = 3 | 0,8110 | 3 | 5,6350 | 6,9483 | - | - | - |
| | | k = 4 | 0,9263 | 4 | 6,1424 | 6,6313 | - | - | |
| | Bietápico | k = 3 | 0,7810 | 3 | 6,2143 | 7,9572 | - | - | |
| | | k = 4 | 1,1107 | 4 | 8,0589 | 7,2558 | - | - | |
| | Jerárquico | k = 3 | 0,8568 | 3 | 5,5242 | 6,4475 | - | - | |
| | | k = 4 | 0,8061 | 4 | 6,2199 | 7,7156 | - | - | |

Tabla 59. Resultados comparativos entre los cinco algoritmos en IRIS

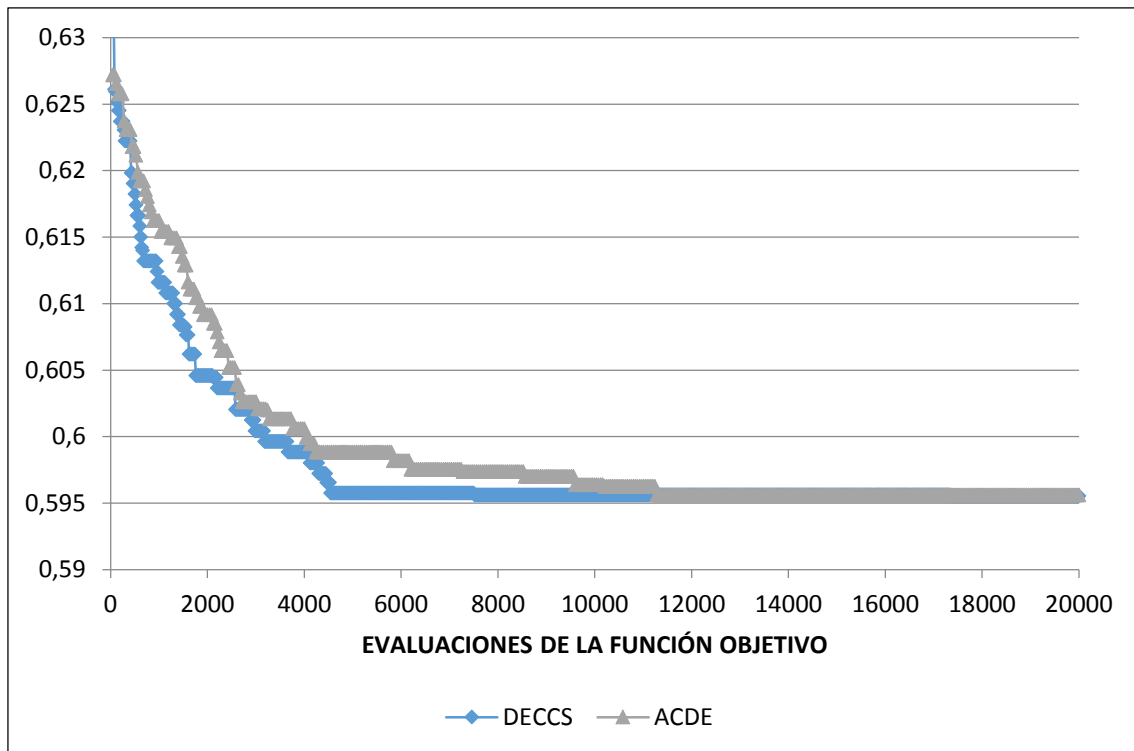


Figura 128. Resultados comparativos entre DECCS y ACDE en IRIS

Problema GLASS

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|--|------------|-----------|-----------------|------|-------------|------------|-------|----------------|---------|
| GLASS Var = 9 Grupos = 6 Vec = 214 | DECCS | Min. | 0,4445 | 3 | 9,3875 | 21,1216 | 40020 | 11,7885 | 1 |
| | | Media | 0,4445 | 3 | 9,3875 | 21,1216 | 40020 | 13,5461 | |
| | | Max. | 0,4445 | 3 | 9,3875 | 21,1216 | 40020 | 14,4566 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0,0000 | 0 | 0,5550 | |
| | ACDE | Min. | 0,4445 | 3 | 9,3875 | 21,1216 | 45090 | 11,2256 | |
| | | Media | 0,4445 | 3 | 9,3875 | 21,1216 | 45090 | 11,4939 | |
| | | Max. | 0,4445 | 3 | 9,3875 | 21,1216 | 45090 | 11,7804 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0,0000 | 0 | 0,1186 | |
| | K-means | k = 3 | 0,9498 | 3 | 14,1443 | 14,8920 | - | - | - |
| | | k = 6 | 0,8613 | 6 | 18,7022 | 21,7150 | - | - | |
| | Bietápico | k = 3 | 0,5856 | 3 | 12,7536 | 21,7810 | - | - | |
| | | k = 6 | 0,5969 | 6 | 11,2304 | 18,8140 | - | - | |
| | Jerárquico | k = 3 | 1,4821 | 3 | 16,2250 | 10,9470 | - | - | |
| | | k = 6 | 0,8027 | 6 | 17,2757 | 21,5210 | - | - | |

Tabla 60. Resultados comparativos entre los cinco algoritmos en GLASS

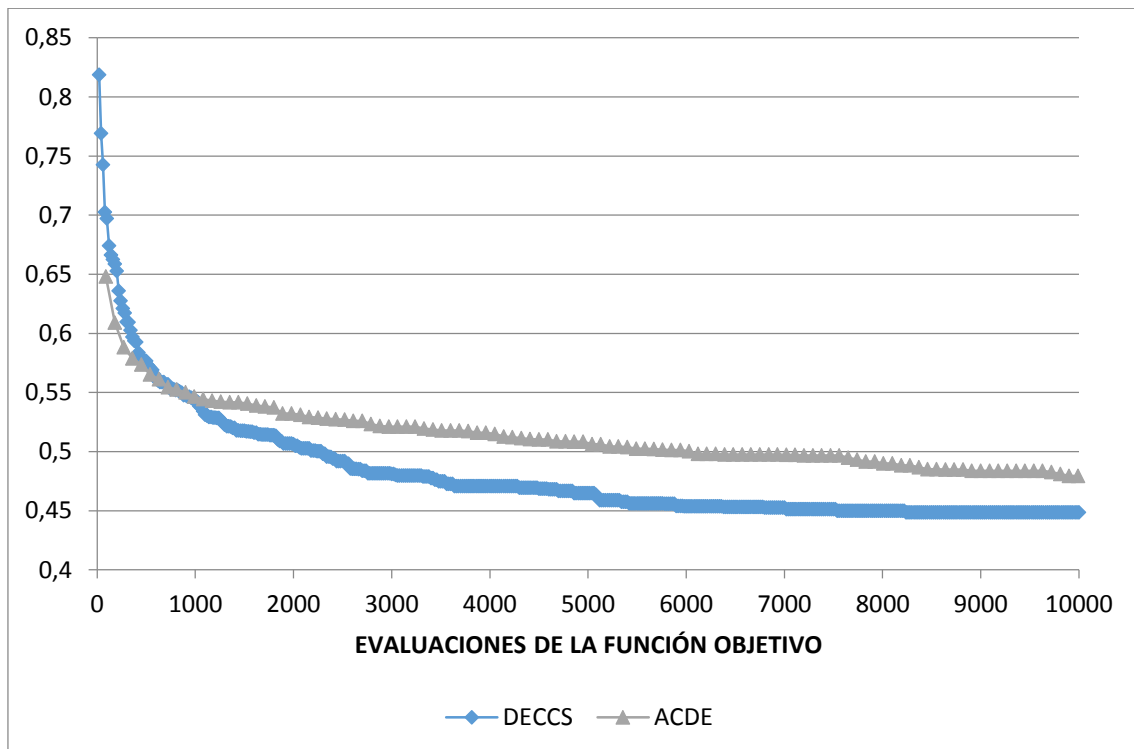


Figura 129. Resultados comparativos entre DECCS y ACDE en GLASS

Problema WINE

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|---|------------|-----------|-----------------|------|-------------|------------|-------|------------|---------|
| WINE Var = 13 Grupos = 3 Vec = 178 | DECCS | Min. | 0,6021 | 2 | 1055,55 | 1752,990 | 26020 | 6,0365 | 1 |
| | | Media | 0,6021 | 2 | 1055,55 | 1752,990 | 26020 | 6,4424 | |
| | | Max. | 0,6021 | 2 | 1055,55 | 1752,990 | 26020 | 6,6500 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0,0000 | 0 | 0,1220 | |
| | ACDE | Min. | 0,6021 | 2 | 1055,55 | 1752,990 | 65130 | 9,7959 | |
| | | Media | 0,6021 | 2 | 1055,55 | 1752,990 | 65130 | 10,0097 | |
| | | Max. | 0,6021 | 2 | 1055,55 | 1752,990 | 65130 | 10,2352 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0,0000 | 0 | 0,1037 | |
| | K-means | k = 2 | 0,8287 | 2 | 995,799 | 1201,640 | - | - | - |
| | | k = 3 | 0,8651 | 3 | 1014,23 | 1172,340 | - | - | |
| | Bietápico | k = 2 | 1,9923 | 2 | 1173,73 | 589,120 | - | - | |
| | | k = 3 | 3,9393 | 3 | 1468,11 | 372,686 | - | - | |
| | Jerárquico | k = 2 | 0,8285 | 2 | 1004,94 | 1212,980 | - | - | |
| | | k = 3 | 0,7278 | 3 | 972,513 | 1336,290 | - | - | |

Tabla 61. Resultados comparativos entre los cinco algoritmos en WINE

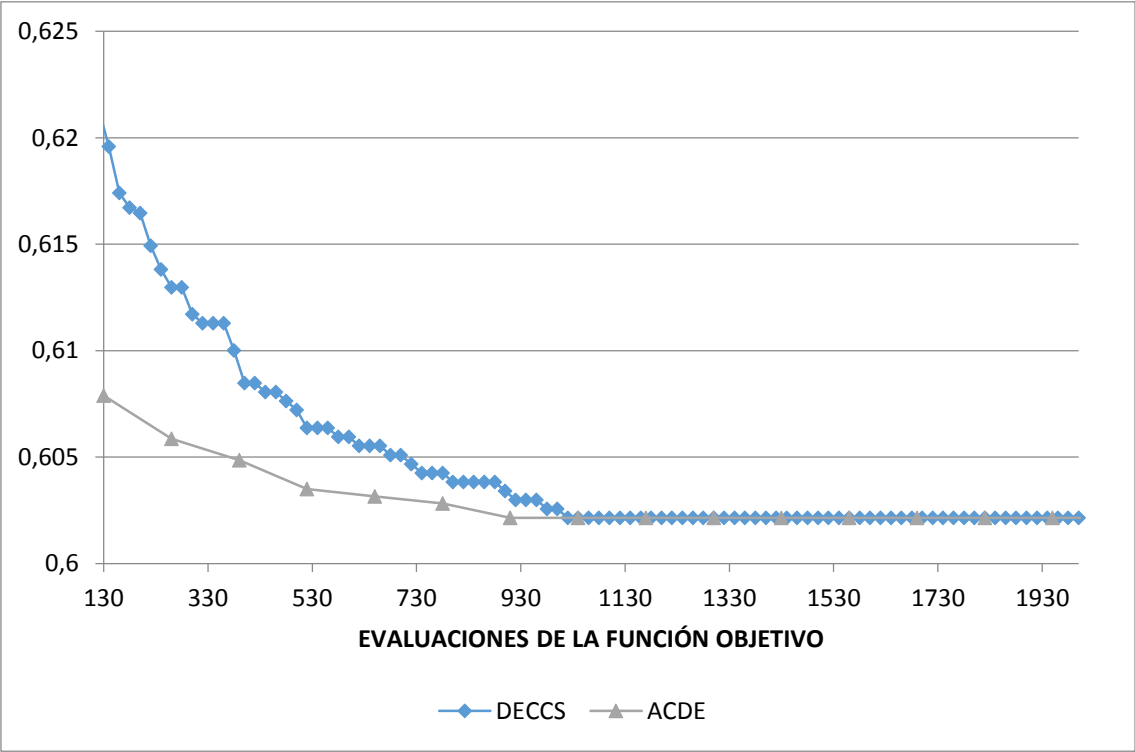


Figura 130. Resultados comparativos entre DECCS y ACDE en WINE

Problema BREAST

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|---|------------|-----------|-----------------|----------|-------------|------------|-------|----------------|---------|
| BREAST Var = 9 Grupos = 2 Vec = 699 | DECCS | Min. | 0,7685 | 2 | 24,1974 | 30,3084 | 44020 | 67,8429 | 0,0000 |
| | | Media | 0,8336 | 2 | 27,9184 | 33,4983 | 44020 | 73,2432 | |
| | | Max. | 0,907 | 2 | 29,7383 | 35,4472 | 44020 | 78,2333 | |
| | | DE | 0,0432 | 0 | 1,7271 | 1,3919 | 0 | 2,5103 | |
| | ACDE | Min. | 0,9826 | 2 | 23,7906 | 24,2111 | 45090 | 80,6096 | |
| | | Media | 1,1447 | 2,0200 | 33,8215 | 29,5654 | 45090 | 81,2779 | |
| | | Max. | 1,3105 | 3 | 40,9526 | 32,6843 | 45090 | 82,9528 | |
| | | DE | 0,0666 | 0,1400 | 2,4865 | 1,7551 | 0 | 0,3476 | |
| | K-means | k = 2 | 1,1012 | 2 | 30,3291 | 27,5431 | - | - | - |
| | | k = 3 | 1,6543 | 3 | 43,1092 | 26,0595 | - | - | |
| | Bietápico | k = 2 | 1,1052 | 2 | 30,3534 | 27,4631 | - | - | |
| | | k = 3 | 1,5063 | 3 | 43,8384 | 29,1035 | - | - | |
| | Jerárquico | k = 2 | 1,1490 | 2 | 34,3777 | 29,9201 | - | - | |
| | | k = 3 | 1,7566 | 3 | 47,9195 | 27,28 | - | - | |

Tabla 62. Resultados comparativos entre los cinco algoritmos en BREAST

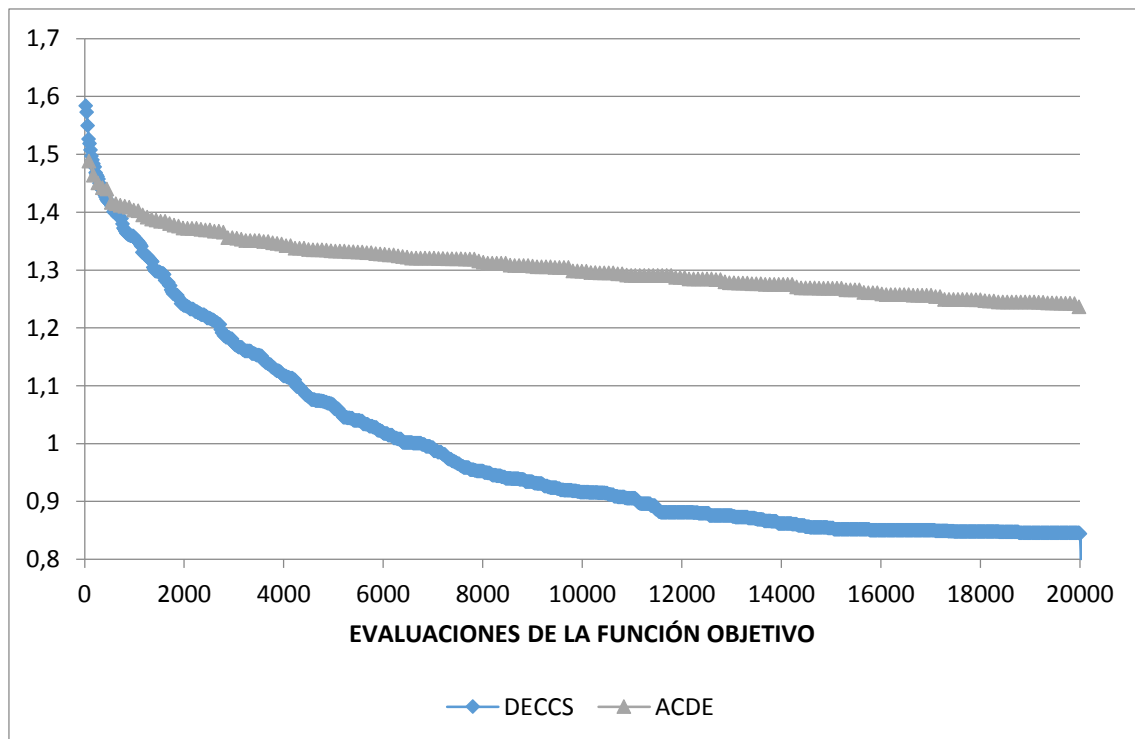


Figura 131. Resultados comparativos entre DECCS y ACDE en BREAST

Problema THYROID

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|--|------------|-----------|-----------------|---------------|-------------|------------|-------|----------------|---------|
| THYROID Var = 5 Grupos = 3 Vec = 215 | DECCS | Min. | 0,4102 | 4 | 80,3406 | 193,8670 | 44020 | 12,4520 | 0,0000 |
| | | Media | 0,4145 | 4,0500 | 80,7859 | 194,8910 | 44020 | 13,3730 | |
| | | Max. | 0,4279 | 5 | 87,9136 | 214,2990 | 44020 | 14,4950 | |
| | | DE | 0,0023 | 0,2200 | 1,6872 | 4,4526 | 0 | 0,5050 | |
| | ACDE | Min. | 0,4144 | 4 | 80,3406 | 191,3960 | 25050 | 5,3820 | |
| | | Media | 0,4286 | 4 | 82,9308 | 193,5160 | 25050 | 5,4550 | |
| | | Max. | 0,4358 | 4 | 84,7457 | 194,5760 | 25050 | 5,5550 | |
| | | DE | 0,0058 | 0 | 1,0205 | 0,8450 | 0 | 0,0390 | |
| | K-means | k = 3 | 1,8265 | 3 | 121,12 | 66,312 | - | - | - |
| | | k = 4 | 1,2088 | 4 | 133,848 | 110,728 | - | - | |
| | Bietápico | k = 3 | 0,9835 | 3 | 113,819 | 115,73 | - | - | |
| | | k = 4 | 0,9663 | 4 | 133,124 | 137,761 | - | - | |
| | Jerárquico | k = 3 | 0,8690 | 3 | 98,1972 | 112,997 | - | - | |
| | | k = 4 | 1,2994 | 4 | 132,496 | 101,965 | - | - | |

Tabla 63. Resultados comparativos entre los cinco algoritmos en THYROID

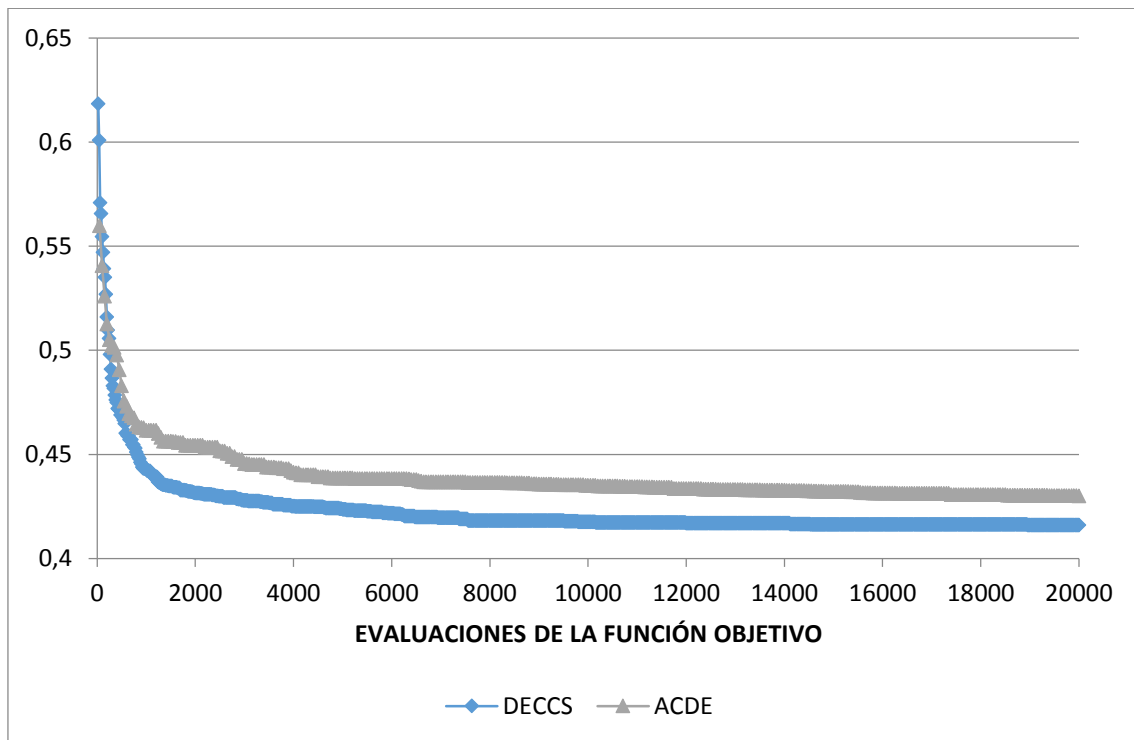


Figura 132. Resultados comparativos entre DECCS y ACDE en THYROID

Problema LED

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|---|------------|-----------|-----------------|---------------|----------------|----------------|--------------|----------------|---------|
| LED Var = 7 Grupos = 10 Vec = 500 | DECCS | Min. | 0,7518 | 12 | 13,6970 | 14,8502 | 30020 | 36,8300 | 0,0000 |
| | | Media | 0,8522 | 16,800 | 16,3579 | 19,2545 | 30020 | 38,2400 | |
| | | Max. | 0,9414 | 19 | 19,2910 | 22,2278 | 30020 | 39,7400 | |
| | | DE | 0,0523 | 1,630 | 1,2859 | 1,7448 | 0 | 0,6700 | |
| | ACDE | Min. | 1,1310 | 13 | 17,1719 | 14,3404 | 35070 | 37,8000 | |
| | | Media | 1,1949 | 15,880 | 20,5601 | 17,2065 | 35070 | 38,2600 | |
| | | Max. | 1,2374 | 20 | 24,0225 | 20,0502 | 35070 | 39,0100 | |
| | | DE | 0,0273 | 1,65 | 1,8128 | 1,4661 | 0 | 0,2700 | |
| | K-means | k = 10 | 1,4051 | 10 | 15,9922 | 11,3813 | - | - | - |
| | | k = 16 | 1,3227 | 16 | 22,3877 | 16,9261 | - | - | |
| | Bietápico | k = 10 | 1,4039 | 10 | 14,3707 | 10,2362 | - | - | |
| | | k = 16 | 1,3145 | 16 | 19,2167 | 14,6187 | - | - | |
| | Jerárquico | k = 10 | 1,5918 | 10 | 16,6110 | 10,4356 | - | - | |
| | | k = 16 | 1,3958 | 16 | 22,5312 | 16,1417 | - | - | |

Tabla 64. Resultados comparativos entre los cinco algoritmos en LED

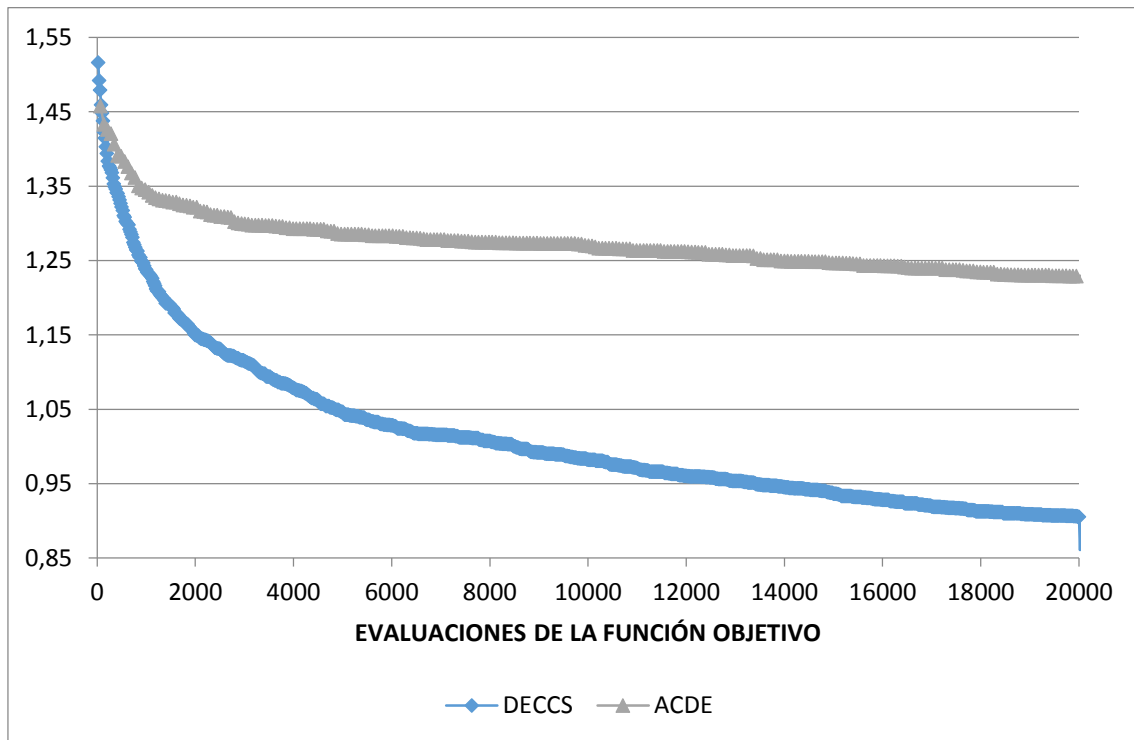


Figura 133. Resultados comparativos entre los cinco algoritmos en LED

Problema YEAST

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|---|------------|-----------|-----------------|---------------|---------------|---------------|-------|-----------------|---------|
| YEAST Var = 8 Grupos = 10 Vec = 1484 | DECCS | Min. | 0,3362 | 4 | 1,2180 | 3,2318 | 40020 | 268,4700 | 0,0000 |
| | | Media | 0,3793 | 5,2800 | 1,5477 | 4,0808 | 40020 | 313,1500 | |
| | | Max. | 0,4147 | 6 | 1,8781 | 4,6224 | 40020 | 326,0700 | |
| | | DE | 0,0216 | 0,7100 | 0,2008 | 0,4787 | 0 | 10,0900 | |
| | ACDE | Min. | 0,4397 | 2 | 0,9636 | 1,8514 | 40080 | 335,8200 | |
| | | Media | 0,4870 | 3,5800 | 1,4076 | 2,8976 | 40080 | 338,9900 | |
| | | Max. | 0,5205 | 5 | 1,9810 | 3,8774 | 40080 | 341,0000 | |
| | | DE | 0,0225 | 0,6300 | 0,1883 | 0,4124 | 0 | 1,2500 | |
| | K-means | k = 5 | 1,6832 | 5 | 3,2765 | 1,9466 | - | - | - |
| | | k = 10 | 1,2729 | 10 | 4,9099 | 3,8573 | - | - | |
| | Bietápico | k = 5 | 1,2136 | 5 | 3,4482 | 2,8412 | - | - | |
| | | k = 10 | 1,1991 | 10 | 4,4481 | 3,7096 | - | - | |
| | Jerárquico | k = 5 | 1,5040 | 5 | 3,1230 | 2,0764 | - | - | |
| | | k = 10 | 1,7227 | 10 | 5,5581 | 3,2264 | - | - | |

Tabla 65. Resultados comparativos entre los cinco algoritmos en YEAST

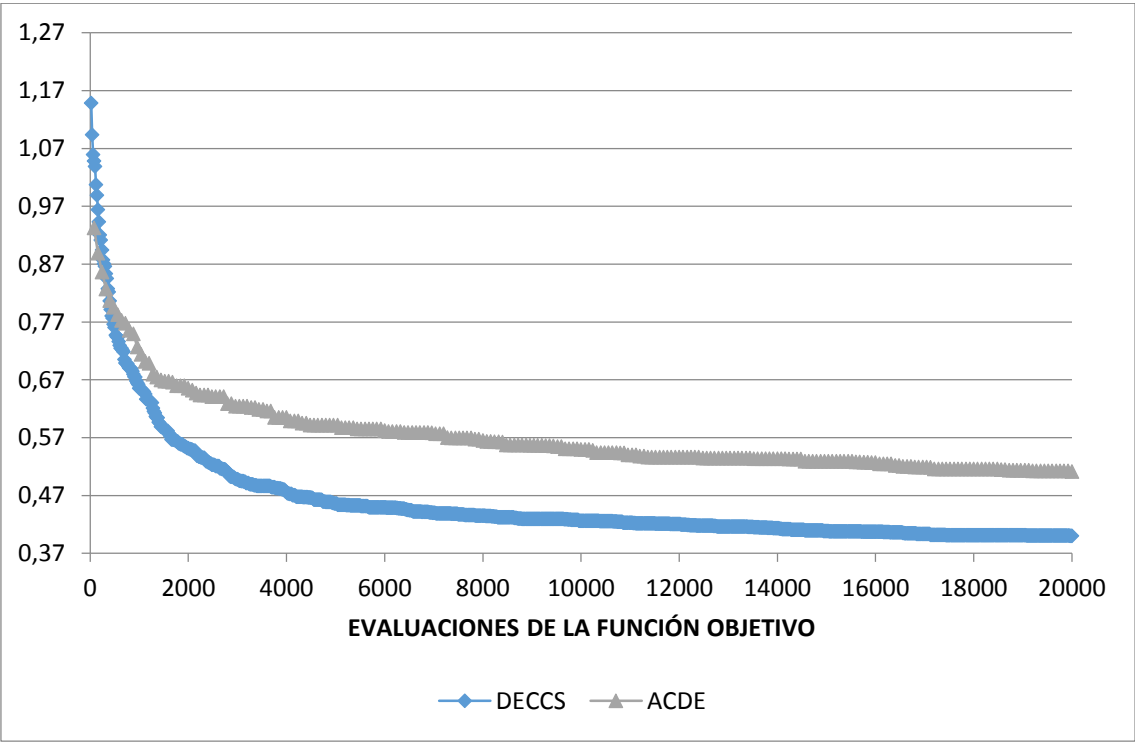


Figura 134. Resultados comparativos entre DECCS y ACDE en YEAST

Problema WINE QUALITY

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|---|------------|-----------|-----------------|----------|-------------|------------|-------|-----------------|---------|
| WINE QUALITY Var = 11 Grupos = 6 Vec = 1599 | DECCS | Min. | 0,3026 | 2 | 144,2790 | 476,7860 | 22020 | 171,9900 | 1 |
| | | Media | 0,3026 | 2 | 144,2790 | 476,7860 | 22020 | 176,2600 | |
| | | Max. | 0,3026 | 2 | 144,2790 | 476,7860 | 22020 | 183,2400 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0,0000 | 0 | 2,0400 | |
| | ACDE | Min. | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 185,3500 | |
| | | Media | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 187,5000 | |
| | | Max. | 0,3026 | 2 | 144,2790 | 476,7860 | 22110 | 190,9500 | |
| | | DE | 0,0000 | 0 | 0,0000 | 0,0000 | 0 | 1,3000 | |
| | K-means | k = 2 | 0,6370 | 2 | 253,8330 | 398,477 | - | - | - |
| | | k = 6 | 0,6569 | 6 | 192,2880 | 292,71 | - | - | |
| | Bietápico | k = 2 | 8,2058 | 2 | 272,6180 | 33,223 | - | - | |
| | | k = 6 | 1,5866 | 6 | 408,9280 | 257,742 | - | - | |
| | Jerárquico | k = 2 | 2,4367 | 2 | 330,2210 | 135,519 | - | - | |
| | | k = 6 | 1,1357 | 6 | 322,6130 | 284,061 | - | - | |

Tabla 66. Resultados comparativos entre los cinco algoritmos en WINE QUALITY

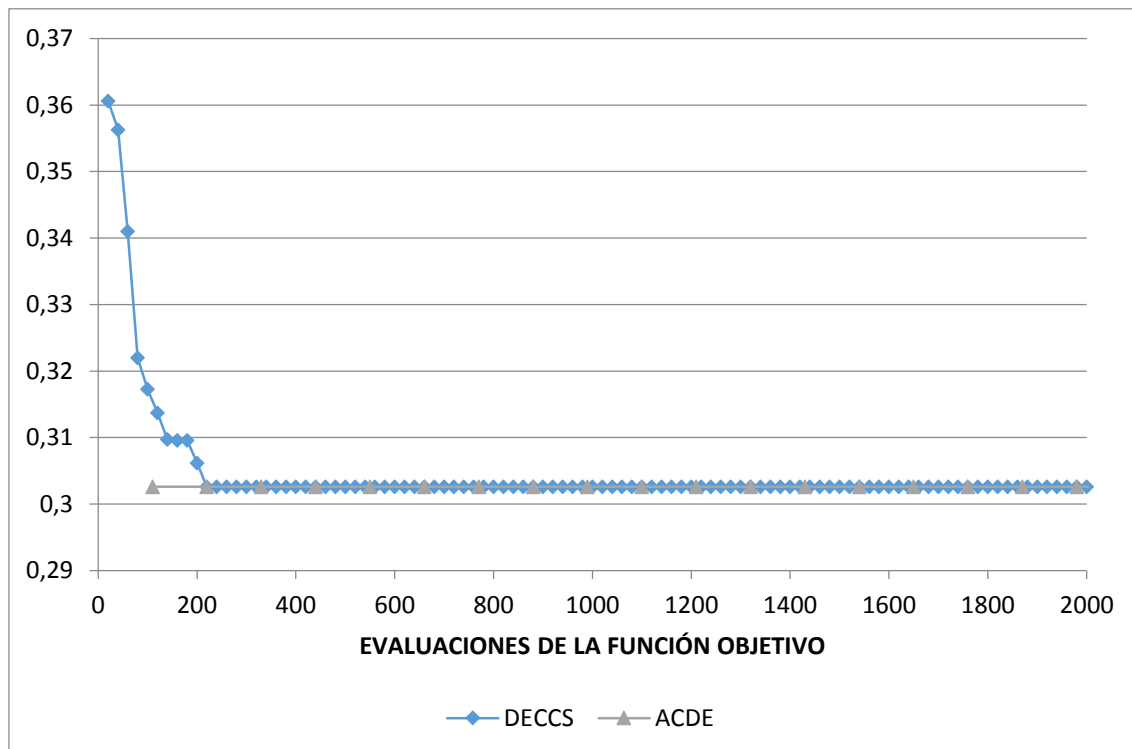


Figura 135. Resultados comparativos entre DECCS y ACDE en WINE QUALITY

Problema SONAR

| Conjunto de datos | Algoritmo | Indicador | Func. Obj. (CS) | # cl | Dist. Intra | Dist Inter | Eval | Tiempo (s) | P-valor |
|---|------------|-----------|-----------------|---------------|---------------|---------------|--------|----------------|---------|
| SONAR Var = 60 Grupos = 2 Vec = 208 | DECCS | Min. | 0,6194 | 2 | 3,1898 | 3,9832 | 50020 | 21,2000 | 0,0000 |
| | | Media | 0,7281 | 2,9500 | 4,0382 | 5,5983 | 50020 | 22,7400 | |
| | | Max. | 0,8150 | 4 | 5,2035 | 7,4370 | 50020 | 24,4900 | |
| | | DE | 0,0505 | 0,5890 | 0,5286 | 0,9601 | 0 | 0,7700 | |
| | ACDE | Min. | 0,7419 | 2 | 3,1898 | 3,9587 | 600600 | 220,0200 | |
| | | Media | 0,7981 | 2,4250 | 3,7501 | 4,7071 | 600600 | 221,4500 | |
| | | Max. | 0,8281 | 3 | 4,6768 | 5,8572 | 600600 | 223,0800 | |
| | | DE | 0,0161 | 0,4940 | 0,6331 | 0,8389 | 0 | 0,6000 | |
| | K-means | k = 2 | 1,8803 | 2 | 4,7561 | 2,5294 | - | - | - |
| | | k = 3 | 1,7556 | 3 | 6,3466 | 3,6150 | - | - | |
| | Bietápico | k = 2 | 2,5946 | 2 | 4,5133 | 1,7395 | - | - | |
| | | k = 3 | 2,0630 | 3 | 5,0239 | 2,4352 | - | - | |
| | Jerárquico | k = 2 | 2,0090 | 2 | 4,9202 | 2,4491 | - | - | |
| | | k = 3 | 1,7962 | 3 | 6,6299 | 3,6911 | - | - | |

Tabla 67. Resultados comparativos entre los cinco algoritmos en SONAR

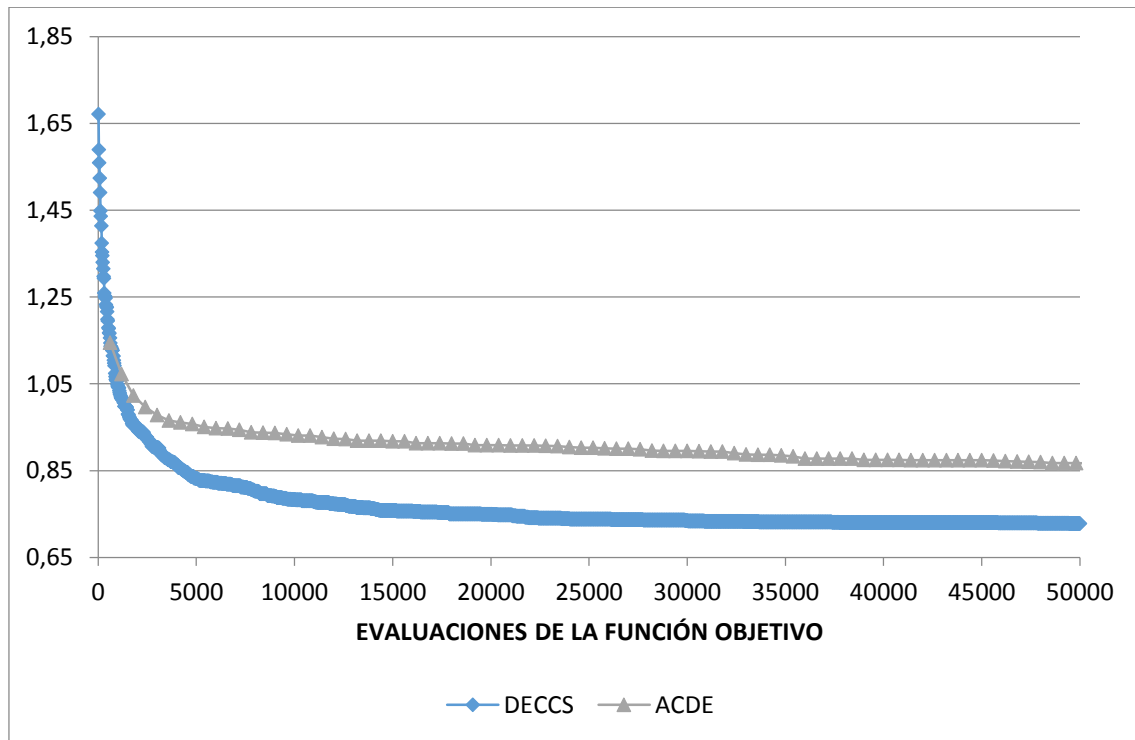


Figura 136. Resultados comparativos entre DECCS y ACDE en SONAR

3.5.2 Velocidades y tiempos de convergencia de DECCS

Otro aspecto a destacar de DECCS, ya comentado en diversas ocasiones en esta investigación, es acerca de los beneficios de su uso en problemas de gran tamaño. Acabamos de mostrar, no solo que obtiene mejores soluciones en la mayoría de los casos, sino que su velocidad de convergencia, y por tanto su capacidad de exploración del espacio de soluciones, es mucho mayor.

En esta sección, vamos a cuantificar dichos beneficios. En primer lugar, comparamos los tiempos totales de ACDE y DECCS. Como se puede apreciar en la Tabla 68, DECCS requiere más tiempo para computar el mismo número de soluciones en *IRIS*, *GLASS*, *WINE* y *THYROID* pero menos en *BREAST* y los cuatro de gran tamaño (*LED*, *YEAST*, *WINEQUALITY*, y *SONAR*). Por tanto, si nos preguntamos cuándo merece la pena el uso de DECCS frente a otros algoritmos, la respuesta es clara y a favor de los problemas de mayor tamaño y complejidad.

| Problema | Variables | Grupos | Vectores | Variables codificadas | Evaluación de la función de aptitud | ACDE | | DECCS | | Diferencia |
|-------------|-----------|--------|----------|-----------------------|-------------------------------------|------------------------|--------|------------------------|--------|------------|
| | | | | | | Tamaño de la población | T(s) | Tamaño de la población | T(s) | |
| IRIS | 4 | 3 | 150 | 30 | 20.000 | 40 | 2,00 | 20 | 3,89 | 94,50% |
| GLASS | 9 | 6 | 214 | 120 | 45.000 | 90 | 11,56 | 20 | 13,55 | 17,21% |
| WINE | 13 | 3 | 178 | 84 | 26.000 | 130 | 3,90 | 20 | 6,44 | 65,13% |
| BREAST | 9 | 2 | 699 | 40 | 45.000 | 90 | 81,80 | 20 | 73,24 | -10,46% |
| THYROID | 5 | 3 | 215 | 36 | 25.000 | 50 | 5,64 | 20 | 7,59 | 34,57% |
| LED | 7 | 10 | 500 | 160 | 30.000 | 70 | 41,36 | 20 | 38,24 | -7,54% |
| YEAST | 8 | 10 | 1.484 | 180 | 40.000 | 80 | 342,48 | 20 | 313,15 | -8,56% |
| WINEQUALITY | 11 | 6 | 1.599 | 144 | 22.000 | 110 | 184,78 | 20 | 176,26 | -4,61% |
| SONAR | 60 | 2 | 208 | 244 | 90.000 | 600 | 34,47 | 20 | 22,74 | -34,03% |

Tabla 68. Comparación entre ACDE y DECCS

Ahora bien, la información que no incluye la tabla anterior es el tiempo que tardaron los algoritmos en alcanzar los óptimos. Podría ocurrir que DECCS consumiese más tiempo que ACDE en el problema de IRIS pero ¿y si encontró el óptimo antes que ACDE y el resto del tiempo lo malgastó?

En la Tabla 69 se ha realizado este experimento. Se selecciona un umbral de parada para ambos algoritmos cercano al óptimo, generalmente entre un 2% y un 5% por encima del óptimo, y se miden las evaluaciones requeridas por ambos algoritmos hasta alcanzarlo.

De nuevo, no nos fijamos en una sola ejecución sino en la media de 40 ejecuciones. Los resultados son contundentes a favor de DECCS pues se obtiene un ahorro medio del 94,23% frente a ACDE.

| Problema | Variables | Grupos | Vectores | Variables codificadas | Evaluaciones requeridas | | Dif. |
|------------|-----------|--------|----------|-----------------------|-------------------------|--------------|----------------|
| | | | | | ACDE | DECCS | |
| IRIS | 4 | 3 | 150 | 30 | 2.240 | 1.620 | -27,68% |
| GLASS | 9 | 6 | 214 | 120 | 22.590 | 6.380 | -71,76% |
| WINE | 13 | 3 | 178 | 84 | 130 | 240 | 84,62% |
| BREAST | 9 | 2 | 699 | 40 | 45.000 | 3.600 | -92,00% |
| THYROID | 5 | 3 | 215 | 36 | 25.000 | 2.940 | -88,24% |
| LED | 7 | 10 | 500 | 160 | 35.000 | 1.400 | -96,00% |
| YEAST | 8 | 10 | 1.484 | 180 | 40.000 | 3.400 | -91,50% |
| VINEQUALIT | 11 | 6 | 1.599 | 144 | 110 | 200 | 81,82% |
| SONAR | 60 | 2 | 208 | 244 | 298.200 | 7.260 | -97,57% |
| | | | | | | Media | -94,23% |

Tabla 69. Comparación entre ACDE y DECCS con umbral de parada

3.5.3 Resultados en bases de datos muy grandes

Aunque hasta ahora nos hemos referido a los problemas *LED*, *YEAST*, *WINE QUALITY* y *SONAR*, como problemas de gran tamaño, obviamente existen otros aún mayores. En esta sección hemos seleccionado uno de ellos, *Pen-Based Recognition of Handwritten Digits data set* (16 variables, 10 grupos y 10.992 vectores), cuya descripción se puede encontrar en la sección 0.

Puesto que el tiempo que requiere cada una de las evaluaciones de una solución es alto, se ha optado por resolverlo únicamente 5 veces con un criterio de parada basado en el tiempo. Tanto ACDE como DECCS se han interrumpido tras una hora de exploración para cada ejecución.

En la Tabla 70 se muestran los resultados desglosados por ejecución para el algoritmo ACDE y en la Tabla 71 los resultados para el algoritmo DECCS. No ha sido posible la realización de un contraste de medias al carecer de suficientes datos. No obstante, se aprecia como el rendimiento medio de DECCS es claramente superior: tanto el valor de medio de DECCS (1,0622) como su desviación estándar (0,0598) son muy superiores a

ACDE (1,5501 de media, con una desviación de 0,0966). Además, la mejor de las 10 ejecuciones la ha obtenido DECCS en su tercera iteración.

| ACDE | | | | | | |
|-----------|---------|-------|---------|---------|--------------|------------|
| Ejecución | FO (CS) | # cl | D.Intra | D.Inter | Evaluaciones | Tiempo (s) |
| 1 | 1,4360 | 4 | 734,36 | 511,38 | 9.280 | 3.608 |
| 2 | 1,5854 | 6 | 1119,27 | 705,97 | 9.440 | 3.656 |
| 3 | 1,4343 | 5 | 868,71 | 605,66 | 9.440 | 3.636 |
| 4 | 1,6391 | 8 | 1449,28 | 884,18 | 9.440 | 3.635 |
| 5 | 1,6555 | 7 | 1218,87 | 736,25 | 9.440 | 3.646 |
| Resumen | | | | | | |
| Min | 1,4343 | 4 | 734,36 | 511,38 | 9.280 | 3.608 |
| Media | 1,5501 | 6 | 1078,10 | 688,69 | 9.408 | 3.636 |
| Max | 1,6555 | 8 | 1449,28 | 884,18 | 9.440 | 3.656 |
| DE | 0,0966 | 1,414 | 253,49 | 125,79 | 64 | 16 |

Tabla 70. Resultados para ACDE en el conjunto Pen-Based Recognition of Handwritten Digits

| DECCS | | | | | | |
|---------|---------|-------|---------|---------|-------|--------|
| Ejec. | FO (CS) | # cl | D.Intra | D.Entre | Eval | Tiempo |
| 1 | 1,0522 | 10 | 1344,43 | 1277,76 | 9.000 | 3.604 |
| 2 | 1,1341 | 8 | 1129,49 | 995,94 | 9.040 | 3.604 |
| 3 | 0,9559 | 7 | 932,23 | 975,24 | 8.920 | 3.601 |
| 4 | 1,0974 | 12 | 1589,99 | 1448,90 | 9.220 | 3.607 |
| 5 | 1,0716 | 7 | 932,45 | 870,18 | 9.040 | 3.604 |
| Resumen | | | | | | |
| Min | 0,9559 | 7 | 932,23 | 870,18 | 8.920 | 3.601 |
| Media | 1,0622 | 8,8 | 1185,72 | 1113,60 | 9.044 | 3.604 |
| Max | 1,1341 | 12 | 1589,99 | 1448,90 | 9.220 | 3.607 |
| DE | 0,0598 | 1,939 | 253,06 | 215,23 | 98 | 2 |

Tabla 71. Resultados para DECCS en el conjunto Pen-Based Recognition of Handwritten Digits

Como ha sido habitual en esta disertación, acompañamos los resultados con una comparativa de la velocidad media de convergencia. En la Figura 137 mostramos la superioridad de DECCS desde las primeras evaluaciones.

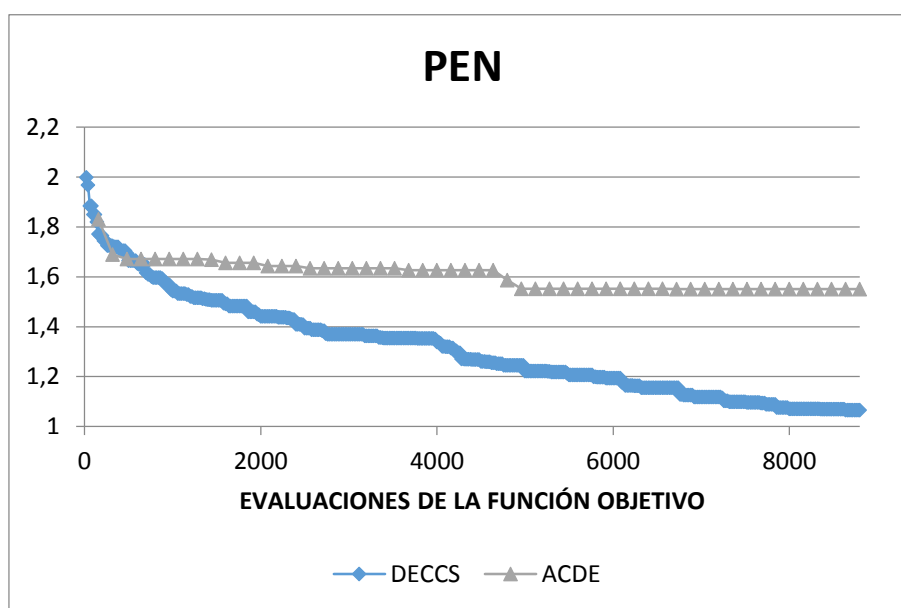


Figura 137. Comparativa de la velocidad media de convergencia entre DECCS Y ACDE

3.5.4 Aplicación a la oficina de farmacia

En la sección 2.2.3 se expuso un conjunto de datos real, con origen en una oficina de farmacia de Sevilla. El problema está claramente motivado por la necesidad de éstas en conocer a sus clientes y así poder adecuar sus estrategias de ventas a los distintos perfiles de sus clientes. Todo ello en línea con lo expuesto en la sección 0.

La base de datos que se ha analizado con el algoritmo DECCS, contiene 13.554 registros y seis variables relacionadas con la venta y el gasto, tanto en productos de venta libre (en la que se incluyen los medicamentos *OTC*), como en medicamentos prescritos con receta médica que son financiados por el sistema sanitario (venta *SOE*).

En esta ocasión, se ha procedido a analizar las soluciones obtenidas por DECCS con las cuatro métricas propuestas, siendo la de *Calinsky-Harabasz* la que ha obtenido unos grupos más compactos y de mayor utilidad para la oficina de farmacia (la métrica CS generó muchos segmentos pequeños frente a uno que contenía al resto).

En la Tabla 72, mostramos los datos más representativos de los dos segmentos obtenidos por DECCS para la farmacia: valor mínimo, máximo, medio y desviación típica de los datos que forman parte del clúster para cada una de las seis variables.

| | | Venta libre | | | Venta SOE | | |
|--|------------|-------------|---------|----------|-----------|----------|----------|
| | | Ventas | PVP | Unidades | Ventas | PVP | Unidades |
| Clúster 1 Tamaño 4762 | Mínimo | 0 | 0 | 0 | 1 | 0,67 | 0 |
| | Media | 0,2974 | 2,6815 | 0,2593 | 2,2585 | 29,5313 | 2,2264 |
| | Máximo | 6 | 246,38 | 6 | 36 | 500,3900 | 19 |
| | Desv. Est. | 0,6746 | 10,9803 | 0,6308 | 1,9832 | 47,1721 | 1,9313 |
| Clúster 2 Tamaño 8780 | Mínimo | 0 | 0 | 0 | 0 | 0 | 0 |
| | Media | 1,5207 | 14,1924 | 1,5369 | 0,0742 | 0,3985 | 0,0711 |
| | Máximo | 22 | 619,56 | 22 | 6 | 93,05 | 6 |
| | Desv. Est. | 1,0443 | 23,4565 | 1,1287 | 0,2932 | 2,9021 | 0,2871 |

Tabla 72. Segmentación obtenida en la aplicación de DECCS en una oficina de farmacia

Un análisis de la composición de los grupos obtenidos por DECCS, se resume en la Figura 138.

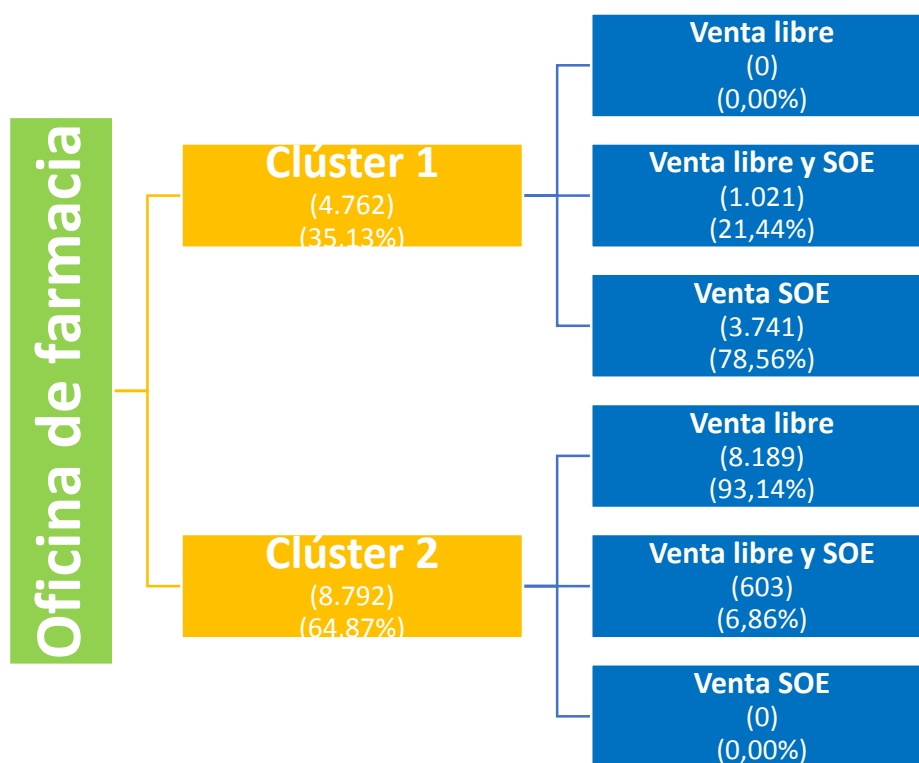


Figura 138. Composición de los grupos obtenidos por DECCS en la oficina de farmacia seleccionada

El reparto de las ventas entre las categorías “venta libre” y “venta SOE”, se muestra con claridad en la Figura 139.

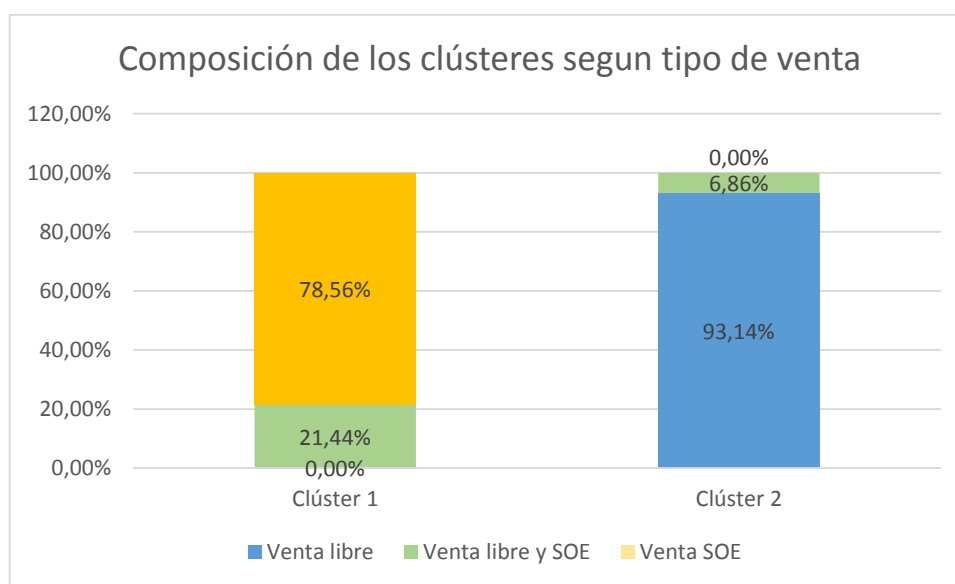


Figura 139. Reparto de las ventas entre las categorías “venta libre” y “venta SOE”

Con todo ello, podemos obtener la caracterización de los dos clústeres obtenidos en el proceso de segmentación:

- **Clúster 1:**

Está formado por 4.762 elementos que suponen el 35,13% del total. En este grupo, es característico que no hay ventas realizadas exclusivamente en la modalidad “venta libre”. Las ventas en este grupo se reparten entre el 78,56%, que son exclusivamente ventas de medicamentos con receta y el 21,44%, en las que se venden productos de los dos tipos.

- **Clúster 2:**

Este grupo está formado por 8.792 elementos que suponen el 64,87% del total. Al contrario que en el clúster 1, este grupo es característico porque no existen ventas realizadas en exclusividad de medicamentos con receta (venta SOE). Las ventas en este grupo se reparten entre el 93,14%, que son exclusivamente ventas de la modalidad venta libre, y sólo un 6,86%, en las que se venden productos de los dos tipos.

A modo de resumen, en el clúster 1 no hay ventas únicamente del tipo libre y el 78,56% son de medicamentos con receta. En el caso del clúster 2, no hay ventas únicamente de medicamentos con receta, pero a diferencia del clúster 1, el porcentaje de ventas que son exclusivamente de la modalidad “venta libre” sube hasta el 93,14%. Es decir, en el clúster 2 la división entre los usuarios que acuden a adquirir solo productos de venta libre es mucho más dominante que lo es la venta SOE en exclusiva en el clúster 1. Esta diferencia queda en un orden de magnitud de tres a uno a favor del clúster 2.

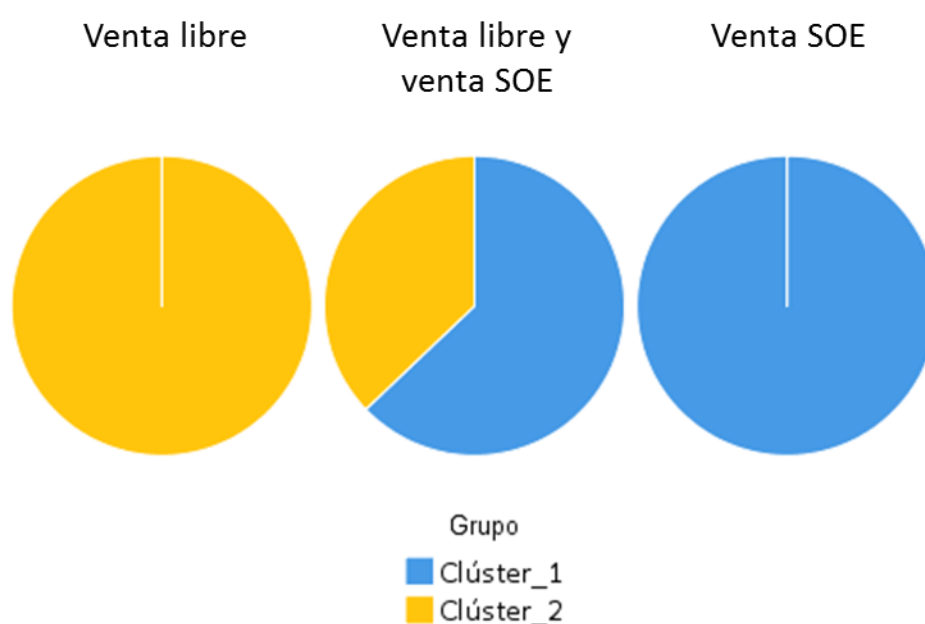


Figura 140. Tipo de venta según clústeres

Conclusiones

El algoritmo DECCS ha identificado dos clústeres. En el primero, predomina la venta de medicamentos con receta (venta SOE) y no existen ventas en las que solo se sirvieron productos de venta libre. En el segundo clúster, sin embargo, predominan las ventas de tipo libre, no existiendo ventas en las que sólo se dispensaron medicamentos con receta. Además, se llega a una conclusión llamativa y es que en esta farmacia en el 2014, existieron dos grandes grupos: uno, que representa el 64,87% de todas las ventas, en el que los clientes que fueron a comprar productos de venta libre, apenas retiraron medicamentos con receta. Y otro, que representa el 35,13% restante, en el que los

pacientes que fueron a retirar los medicamentos prescritos por el médico, también compraron productos de venta libre en un 21,44% de las ocasiones.

En cuanto esté terminada la integración del programa de gestión de la oficina de farmacia con la aplicación de clientes, tarea que se está llevando por otra investigación, se podrá obtener mayor información de los grupos y, por tanto, la farmacia podrá diseñar una estrategia que permita tratar a los dos grupos de clientes obtenidos con mayor personalización.

3.6 Resultados de la encuesta a farmacéuticos de la provincia de Sevilla

En Sevilla y su provincia existen 786 oficinas de farmacia según el Consejo General de Colegios Oficiales de Farmacéuticos (31). Dado que se recibieron 183 cuestionarios, la tasa de respuesta fue del 23,28%. Se considera esta como una cifra relevante, al tener en cuenta que, según el propio Colegio, la tasa de respuesta habitual se sitúa en torno al 12%. La propia situación del sector y la promoción de la venta de medicamentos sin receta por internet en el verano de 2014 por parte de la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) son hechos bastante notorios como para estimular al conjunto de farmacéuticos comunitarios a permanecer alerta ante nuevas formas de desarrollo de su negocio.

Como ya se adelantó en la sección 2.4, la intención de este estudio realizado a las oficinas de farmacia de Sevilla, no tiene el objetivo de estimar parámetros concretos, sino más bien, obtener una aproximación del interés e inquietud por nuevos planteamientos en su negocio y por la incorporación del término “cliente” en su vocabulario habitual.

A continuación, se presentan los principales resultados agrupados en tres bloques. Estos se corresponden con la caracterización de la oficina de farmacia, con la consideración de los términos “cliente” y “paciente” y con la comunicación con ellos (véase el Anexo 6.3).

LA FARMACIA

Como se observa en la Tabla 73, la oficina de farmacia que ha participado en la encuesta ha sido mayormente urbana, con un 77,60% frente a un 22,40% rural.

| Tipo | Rural | Urbana | Total |
|------|--------|--------|------------|
| Nº | 41 | 142 | 183 |
| % | 22,40% | 77,60% | |

Tabla 73. Pregunta 1: su tipo de farmacia es...

En el caso de que la farmacia sea o no la única en su localidad, Tabla 74 y Tabla 75, la participación de las farmacias que comparten localidad con otras farmacias, es abrumadora con un 96,72%. Si consideramos las farmacias urbanas que no son únicas en su localidad, el porcentaje supone un 77,05% del total de las farmacias participantes.

| Tipo | ¿Única? | | Total |
|--------------|------------|----------|------------|
| | No | Sí | |
| Rural | 36 | 5 | 41 |
| Urbana | 141 | 1 | 142 |
| Total | 177 | 6 | 183 |

Tabla 74. Preguntas 1 y 2: tipo de farmacia según si es la única o no en el municipio. Frecuencias absolutas

| Tipo | ¿Única? | | Total |
|--------------|---------------|--------------|----------------|
| | No | Sí | |
| Rural | 19,67% | 2,73% | 22,40% |
| Urbana | 77,05% | 0,55% | 77,60% |
| Total | 96,72% | 3,28% | 100,00% |

Tabla 75. Preguntas 1 y 2: tipo de farmacia según si es la única o no en el municipio. Porcentajes sobre el total

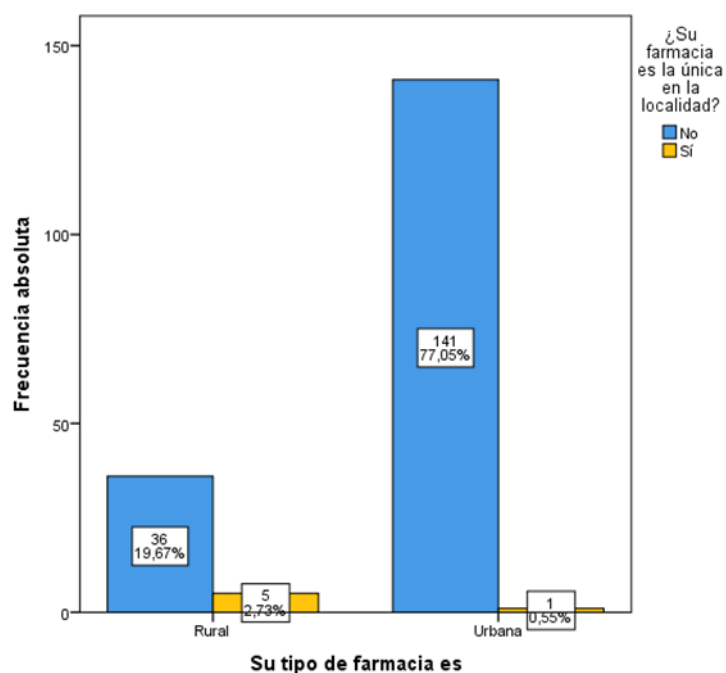


Figura 141. Tipo de farmacia y presencia única o no en la localidad

Para ahondar más en el perfil de la farmacia participante, se ha cruzado el tipo de farmacia y si su distancia al centro de salud más próximo es inferior a 250 metros. Las frecuencias absolutas y los porcentajes sobre el total, aparecen en la Tabla 76 y en la Tabla 77, respectivamente. En ellas, se muestra que la mayoría de las farmacias participantes, el 61,75%, son urbanas con una distancia superior a 250 metros al centro de salud más cercano. Porcentaje que sube hasta el 81,97% si se incluyen las farmacias rurales.

| Tipo | ¿Distancia < 250 m? | | Total |
|--------|---------------------|----|-------|
| | No | Sí | |
| Rural | 37 | 4 | 41 |
| Urbana | 113 | 29 | 142 |
| Total | 177 | 6 | 183 |

Tabla 76. Preguntas 1 y 3: tipo de farmacia según si la distancia al centro de salud más próximo es inferior a 250 metros. Frecuencias absolutas

| Tipo | ¿Distancia < 250 m? | | Total |
|--------------|---------------------|---------------|----------------|
| | No | Sí | |
| Rural | 20,22% | 2,19% | 22,40% |
| Urbana | 61,75% | 15,85% | 77,60% |
| Total | 81,97% | 18,03% | 100,00% |

Tabla 77. Preguntas 1 y 3: tipo de farmacia según si la distancia al centro de salud más próximo es inferior a 250 metros. Porcentajes sobre el total

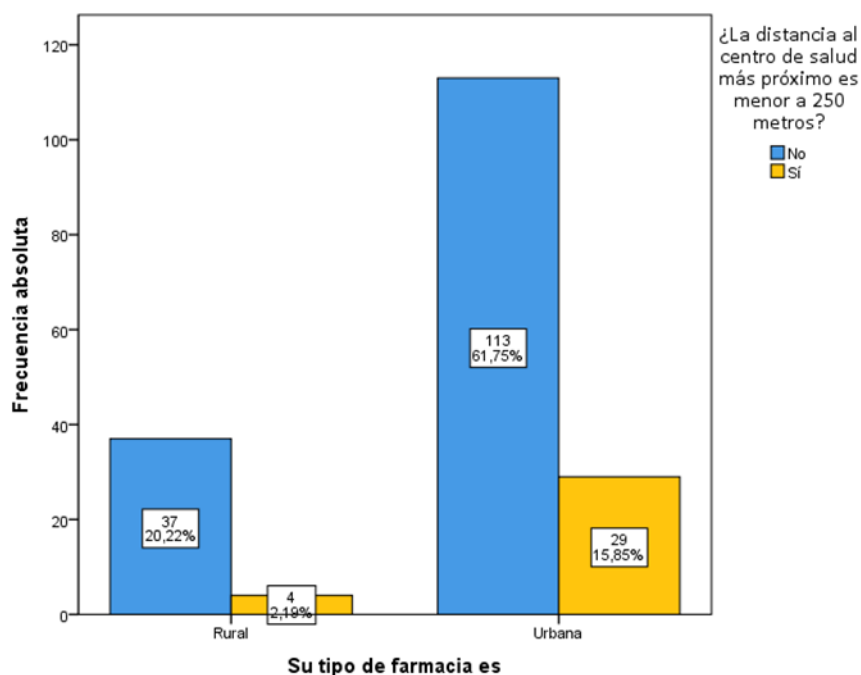


Figura 142. Tipo de farmacia según si la distancia al centro de salud más próximo es inferior a 250 metros

Como conclusión, en el apartado de caracterización de la oficina participante, podemos afirmar que se trata principalmente de una oficina de farmacia urbana, que no es única en su municipio y cuya distancia al centro de salud más próximo es superior a 250 metros. Esto ha ocurrido en 112 ocasiones lo que supone un 61,20%. Esta característica confirma la inquietud de las oficinas de farmacia respecto a la situación actual de las mismas, puesto que son farmacias que sienten la necesidad de nuevos planteamientos en mayor o menor medida.

COMUNICACIÓN CON CLIENTES/PACIENTES

En este bloque, analizamos el grado en que los términos “cliente” y “paciente” se han incorporado a su práctica habitual.

Resulta llamativo, analizando la Tabla 78 y la Tabla 79, como el 59,02% de las oficinas de farmacia participantes sí distinguen los términos “cliente” y “paciente”. Sin embargo, sólo el 26,23% del total de participantes utiliza una tarjeta de fidelización con sus clientes.

| Nº | ¿Tarjeta de fidelización? | | Total |
|------------------------|---------------------------|-----------|------------|
| | No | Sí | |
| ¿Clientes / pacientes? | No | Sí | |
| No | 44 | 31 | 75 |
| Sí | 60 | 48 | 108 |
| Total | 104 | 79 | 183 |

Tabla 78. Preguntas 4 y 7: distinción clientes y pacientes según si la oficina de farmacia posee o no tarjeta de fidelización. Frecuencias absolutas

| % | ¿Tarjeta de fidelización? | | Total |
|------------------------|---------------------------|---------------|----------------|
| | No | Sí | |
| ¿Clientes / pacientes? | No | Sí | |
| No | 24,04% | 16,94% | 40,98% |
| Sí | 32,79% | 26,23% | 59,02% |
| Total | 56,83% | 43,17% | 100,00% |

Tabla 79. Preguntas 4 y 7: distinción clientes y pacientes según si la oficina de farmacia posee o no tarjeta de fidelización. Porcentajes sobre el total

Si añadimos la Tabla 80 y la Tabla 82 a este análisis, se muestra como una mayoría abrumadora de estas oficinas de farmacia, no son únicas en su localidad. De hecho, sólo el 0,55% de participantes sobre el total son únicas en su localidad y distinguen entre “clientes” y “pacientes” y poseen tarjeta de fidelización. Aunque de esta clase de farmacia sólo han participado seis, queda claro que el ser la única farmacia en la

localidad resulta ser un factor discriminante en el grado de sensibilidad a estos nuevos términos.

| Nº | | ¿Tarjeta de fidelización? | | |
|-------------------|------------------------|---------------------------|----|------------|
| ¿Única? | ¿Clientes / pacientes? | No | Sí | Total |
| Farmacia no única | No | 43 | 30 | 73 |
| | Sí | 57 | 47 | 104 |
| | Total | 100 | 77 | 177 |
| Farmacia única | No | 1 | 1 | 2 |
| | Sí | 3 | 1 | 4 |
| | Total | 4 | 2 | 6 |
| Total | | 104 | 79 | 183 |

Tabla 80. Preguntas 2,4 y 7: distinción clientes y pacientes según si la oficina de farmacia posee o no tarjeta de fidelización y si la oficina de farmacia es única o no en su localidad. Frecuencias absolutas

| % | | ¿Tarjeta de fidelización? | | |
|-------------------|------------------------|---------------------------|---------------|---------------|
| ¿Única? | ¿Clientes / pacientes? | No | Sí | Total |
| Farmacia no única | No | 23,50% | 16,39% | 39,89% |
| | Sí | 31,15% | 25,68% | 56,83% |
| | Total | 54,65% | 42,07% | 96,72% |
| Farmacia única | No | 0,55% | 0,55% | 1,10% |
| | Sí | 1,64% | 0,55% | 2,19% |
| | Total | 2,19% | 1,10% | 3,29% |
| Total | | 56,84% | 43,17% | 100% |

Tabla 81. Preguntas 2,4 y 7: distinción clientes y pacientes según si la oficina de farmacia posee o no tarjeta de fidelización y si la oficina de farmacia es única o no en su localidad. Porcentajes sobre el total

Esto lo podemos observar de forma visual en la Figura 143 y en la Figura 144.

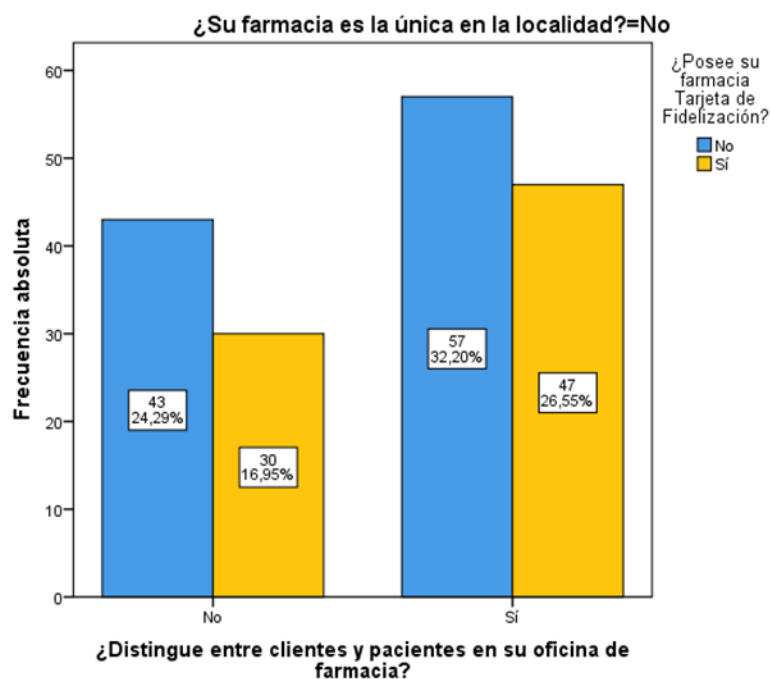


Figura 143. Comparación de la distinción entre la consideración de “clientes” y “pacientes” y si la farmacia posee o no tarjeta de fidelización para farmacias que no son únicas en su localidad

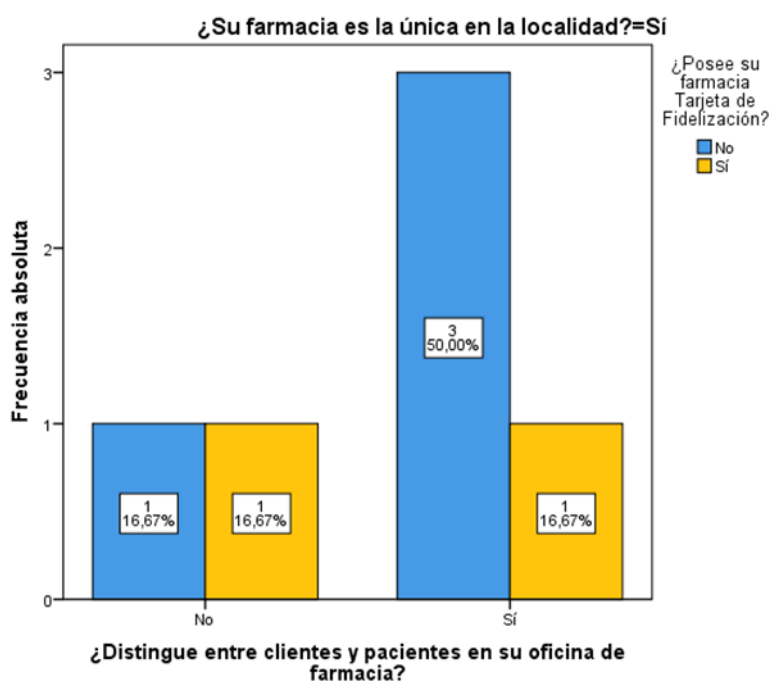


Figura 144. Comparación de la distinción entre la consideración de “clientes” y “pacientes” y si la farmacia posee o no tarjeta de fidelización para farmacias que son únicas en su localidad

CONOCIMIENTO DE CLIENTES/PACIENTES

En este último bloque se va a analizar el conocimiento del que se dispone sobre los clientes de las oficinas de farmacia y sus canales de comunicación.

De la Tabla 82, en la que en verde se destacan los mayores porcentajes, se desprende que el medio de comunicación más utilizado, con un 97,27% sobre el total, es el teléfono. La mayoría de las oficinas participantes declara mayoritariamente, con un 74,32% sobre el total, que no utiliza el correo electrónico con sus clientes. Porcentaje que disminuye hasta el 58,57%, para aquellas que no utilizan el SMS como medio de comunicación. En este caso, cabe destacar que el porcentaje de las que sí lo utilizan es del 41,53% sobre el total. Respecto al uso de las redes sociales, queda de manifiesto que la mayoría no usa ni *Blogs* ni *Twitter* de forma contundente. El caso de Facebook, aunque mayoritariamente no se usa (67,76%), existe un 32,34% que declara usarlo con sus clientes.

| | No | % | Sí | % | Total |
|--------------------------------|-----|--------|-----|--------|------------|
| Correo electrónico | 136 | 74,32% | 47 | 25,68% | 183 |
| SMS (mensajes de móvil) | 107 | 58,47% | 76 | 41,53% | 183 |
| Teléfono | 5 | 2,73% | 178 | 97,27% | 183 |
| Redes sociales | | | | | |
| Blog | 174 | 95,08% | 9 | 4,92% | 183 |
| Twitter | 166 | 90,71% | 17 | 9,29% | 183 |
| Facebook | 124 | 67,76% | 59 | 32,24% | 183 |
| Otros | 122 | 66,67% | 61 | 33,33% | 183 |

Tabla 82. Pregunta 5: ¿Qué medios utiliza para comunicarse con sus clientes/pacientes?

Cuando a las oficinas de farmacia participantes se les pregunta expresamente sobre cuál es el medio de comunicación más usado con sus clientes, los resultados concuerdan con los obtenidos en la pregunta espontánea anterior (Tabla 82).

En este caso, Tabla 83, con un 77,05% sobre el total los participantes se decantan por el teléfono como el medio más utilizado. Está seguido con un 6,56%, por el SMS y Facebook. Y por “otros medios”, con un 3,83%; y por último, el correo electrónico con un 2,73%.

| | Nº | % |
|--------------------|-----|--------|
| Correo electrónico | 5 | 2,73% |
| SMS | 12 | 6,56% |
| Teléfono | 141 | 77,05% |
| Redes sociales | | 0,00% |
| Blog | 0 | 0,00% |
| Twitter | 0 | 0,00% |
| Facebook | 12 | 6,56% |
| Otros | 6 | 3,28% |
| Ninguno | 7 | 3,83% |
| Total | 183 | 100% |

Tabla 83. Pregunta 6: ¿Cuál es el medio que más utiliza?

La representación gráfica de la información de esta tabla, se expone en la Figura 145.

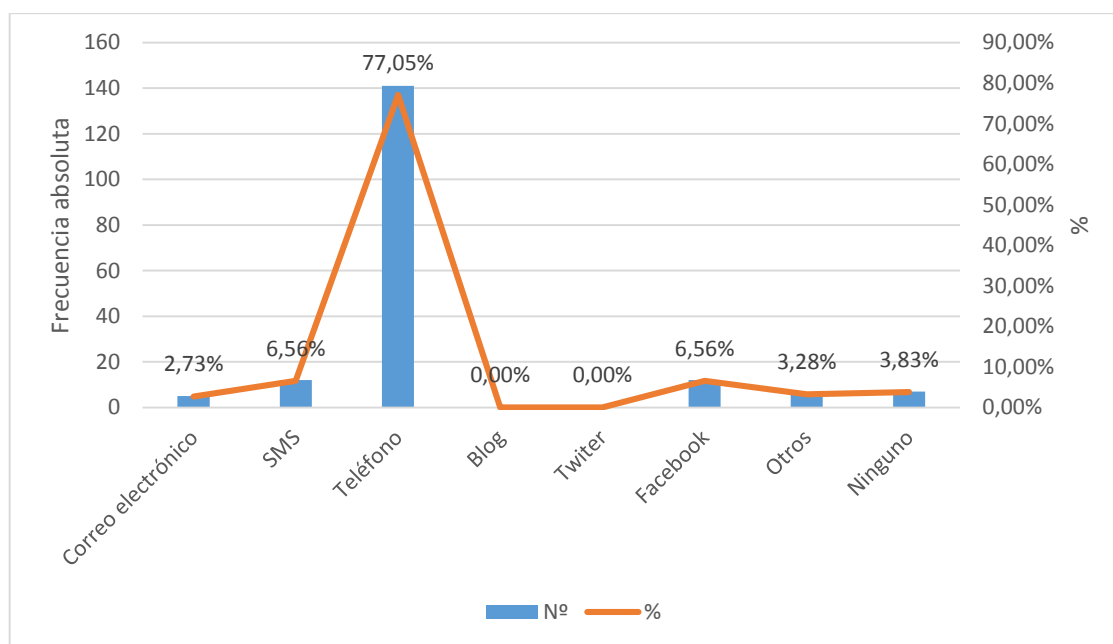


Figura 145. Medio de comunicación más utilizado

Para finalizar este bloque, se analiza el tipo de información que declaran registrar los participantes sobre sus clientes. De la Tabla 84, procediendo con el color verde como en la anterior pregunta, resulta que la mayoría, con un 50,82%, recoge información sobre la frecuencia de la compra o la visita de sus clientes. El 71,58%, sobre el 28,42% que

declara lo contrario, afirma que no recoge información sobre hábitos de consumo. Destaca la recogida de las preferencias de servicios de forma contundente, con un 97,27% sobre el total, siendo esta la información que más se registra de todas las posibles en el estudio. El 61,75% declara no registrar los hábitos para el cuidado de la salud, frente a un 38,25% que declara registrarlos. Por último, aquellas farmacias que declaran registrar los temas de interés para los clientes y otra información relacionada con el conocimiento de sus pacientes, representan el 50,27% y el 56,28%, respectivamente.

| | No | % | Sí | % | Total |
|---|-----|--------|-----|--------|------------|
| Frecuencia de compra/visita | 90 | 49,18% | 93 | 50,82% | 183 |
| Hábitos de consumo (tabaco, alcohol,...) | 131 | 71,58% | 52 | 28,42% | 183 |
| Preferencia en servicios (seguimiento de la medicación, dietas,...) | 5 | 2,73% | 178 | 97,27% | 183 |
| Hábitos para el cuidado de la salud (deporte, control medicación,...) | 113 | 61,75% | 70 | 38,25% | 183 |
| Temas de interés para los pacientes/clientes (dermocosmética, cuidado pediátrico, ...) | 91 | 49,73% | 92 | 50,27% | 183 |
| Otra información relacionada con el conocimiento de los pacientes/clientes | 80 | 43,72% | 103 | 56,28% | 183 |

Tabla 84. Pregunta 7: ¿Qué información registra de sus clientes/pacientes?

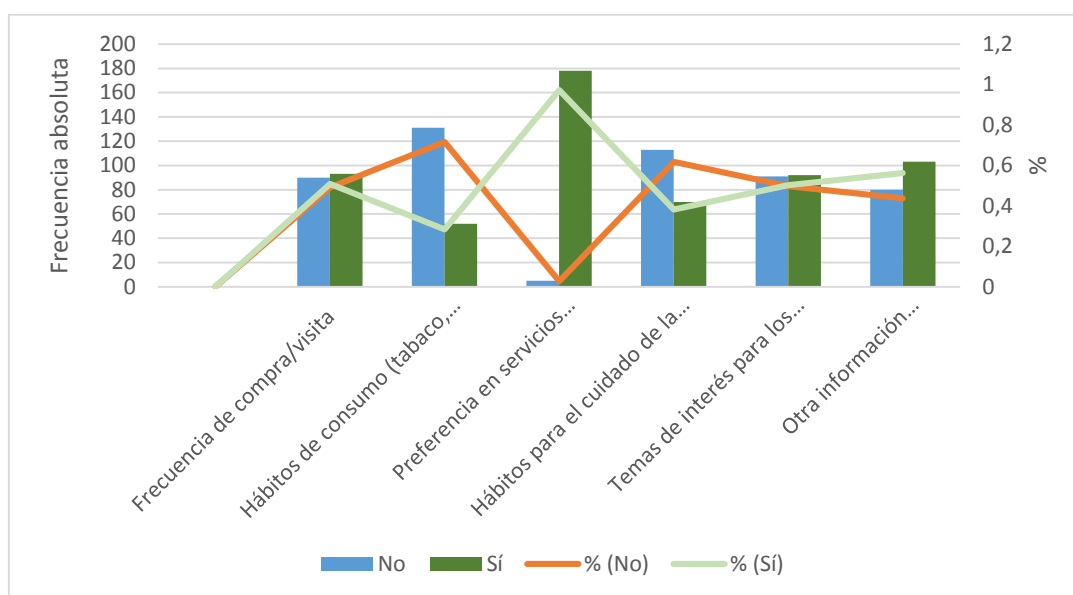


Figura 146. Información registrada de clientes o pacientes

4 RESUMEN Y CONCLUSIONES

4.1 Principales aportaciones

Como ya se expuso en la introducción de esta investigación, el principal objetivo era encontrar un modelo de caracterización de los usuarios (clientes y pacientes) de una oficina de farmacia en base a variables discriminantes y a un algoritmo de segmentación adecuado y contrastado.

Así pues, la aportación principal de esta investigación consiste en el diseño, puesta a punto, confirmación e implementación de un algoritmo bioinspirado de segmentación de los usuarios (clientes y pacientes) en una oficina de farmacia. Este algoritmo es de la familia de los algoritmos metaheurísticos que han sido usados con éxito en multitud de problemas en este campo y en otras áreas de investigación. El algoritmo propuesto, denominado DECCS, mejora al algoritmo tomado como base, llamado ACDE, gracias a la introducción de las siguientes variantes y consideraciones:

- La codificación de las soluciones con un esquema de reordenamiento.
- La utilización de la técnica de hipercubos latinos para la generación de la población inicial.
- La introducción del esquema de continuación, que permite utilizar tamaños de población pequeños acelerando la convergencia sin perder capacidad de exploración.
- Y por último, la utilización de esquemas auto-adaptativos para sus principales parámetros.

En lo que respecta a la situación de la oficina de farmacia en España, la principal aportación queda materializada en la siguiente matriz DAFO:

- Debilidades.
 - Cultura empresarial mejorable por parte del farmacéutico.
 - Bajo uso de herramientas de gestión empresarial.
 - Insuficiente aplicación de las nuevas tecnologías.
 - Alta dependencia del mercado público regulado.
 - Venta libre inferior a lo deseable.

- Baja diferenciación entre farmacias.
- Amenazas:
 - Liberalización del sector.
 - Reducción del gasto público.
 - Medidas de contención de la demanda de medicamentos.
 - Retrasos en los pagos.
 - Salida de medicamentos de la oficina de farmacia hacia la dispensación hospitalaria.
 - Aumento de franquiciados en parafarmacia.
 - Disminución del precio del medicamento financiado.
 - Aparición de cadenas de farmacia.
- Fortalezas:
 - Prestigio de la profesión.
 - Alta valoración de los usuarios.
 - Cercanía y proximidad.
 - Consejo y trato personalizado.
 - Referencia para los pacientes en su salud.
 - Imagen de la farmacia en la población.
- Oportunidades:
 - La población está cada vez más comprometida con el autocuidado de la salud y la prevención.
 - Especialización de las oficinas de farmacia.
 - Aplicación de herramientas de gestión empresarial y marketing.
 - Aplicación de las nuevas tecnologías en el conocimiento de los clientes.
 - Aplicación de técnicas de segmentación y fidelización de clientes.
 - Potencial desarrollo de nuevos servicios remunerados.
 - Mayor aprovechamiento de la receta electrónica.
 - Medicina personalizada.
 - La comunicación médico-farmacéutico.
 - Venta de medicamentos sin receta por internet.

4.2 Conclusiones

- a. Según los objetivos que se pretendan alcanzar, respecto al número de grupos, destacan dos métricas:
 - La métrica CS si deseamos obtener un mayor número de grupos o perfiles de usuarios, pues se muestra más sensible a la hora de crear más grupos de menor tamaño.
 - La métrica CH si desea un menor número de grupos de mayor tamaño.
- b. El esquema de reordenamiento propuesto es claramente beneficioso, en especial, para problemas de gran dimensión puesto que acelera la convergencia sin perder diversidad en la población.
- c. Los resultados obtenidos, en lo que respecta a las técnicas de generación de la población inicial, son contundentes a favor de un muestreo basado en hipercubos latinos frente al aleatorio y al basado en vectores opuestos.
- d. DECCS, el algoritmo bionspirado propuesto, mejora las prestaciones de los algoritmos clásicos seleccionados, tanto en los problemas utilizados de tamaño medio como en los de gran tamaño.
- e. Respecto a la velocidad de convergencia a la solución óptima, DECCS presenta una mayor velocidad de convergencia que ACDE en los cuatro problemas de gran tamaño utilizados.
- f. La aplicación de DECCS a una base de datos del año 2014 de una oficina de farmacia, ha resultado un éxito. El algoritmo ha caracterizado dos grupos claros de clientes según el tipo de producto retirado en la farmacia: medicamento o producto de venta libre (incluyendo medicamentos sin receta “OTC”).
 - El primer grupo se caracteriza por retirar medicamentos y productos de venta libre. No habiendo ventas realizadas exclusivamente de tipo libre.
 - El segundo se caracteriza por retirar productos de venta libre casi de forma exclusiva. En este grupo no hay ventas realizadas exclusivamente de medicamentos.

A modo de resumen, en esta farmacia en el 2014 resultaron dos grandes grupos de usuarios bien diferenciados. El primero de ellos compuesto principalmente por pacientes y el segundo formado claramente por clientes. Además, por la característica del primer grupo, resulta que sobre él se pueden desarrollar prácticas de asesoramiento para que aumente la venta libre, al tratarse de un grupo sensible a ella. Por todo ello, la farmacia podrá diseñar una estrategia diferencial que permita tratar a los dos grupos de usuarios obtenidos con mayor personalización. Resulta evidente, que en esta farmacia la “venta libre” predomina sobre la “venta SOE”, al menos en lo que respecta al año 2014.

- g. Para lograr una explotación máxima de la segmentación obtenida se debe tener en cuenta:
- La inclusión en los programas de gestión de las oficinas de farmacia de la diferenciación entre clientes y pacientes.
 - La actualización periódica del perfil y caracterización de sus usuarios para que refleje la realidad de la oficina de farmacia en cada momento.
- h. Gracias al análisis del sector farmacéutico en España, practicado a sus variables más representativas, queda demostrado que el último lustro ha sido un periodo de mucha tensión. En él, se pone en jaque su modelo, el “modelo mediterráneo”, y por primera vez en mucho tiempo su situación financiera, que se ve deteriorada.
- i. La gran baza de las oficinas de farmacia, la dispensación de medicamentos prescritos con receta, experimenta las consecuencias de la situación económica de España en forma de medidas de constricción de la demanda. Los márgenes bajan tanto para las oficinas de farmacia como para la distribución y la industria farmacéutica.
- j. Del lado de la oferta, entre los años 2011 y 2014, se produce un aumento del consumo de medicamentos genéricos tanto en unidades como en valores, llegando al 38% sobre el total de prescripciones. Hecho que provoca una bajada del ticket medio en las oficinas de farmacia debido al menor precio de éstos frente a los medicamentos de marca.

- k. El gasto farmacéutico público, compuesto por el gasto farmacéutico público en recetas y el gasto farmacéutico público hospitalario, ha sufrido un descenso medio acumulado del 4,16% en el periodo 2008-2012, perdiendo 2,29 puntos de peso sobre el gasto sanitario nacional total en España en dicho periodo. No obstante, el gasto farmacéutico público hospitalario ha experimentado un aumento en dicho periodo de entre el 2,49% y el 4,18% según la fuente consultada. Esta subida ha sido provocada, entre otros, por la desviación de medicamentos más caros desde la oficina de la farmacia, por parte de la industria de medicamentos, a los hospitales. Esto es debido a la falta de “transparencia” del gasto farmacéutico en hospitales, que se constata con la no publicación de la “Estadística de Gasto Farmacéutico de los Hospitales Públicos del Sistema Nacional de Salud” que está incluida en el inventario de operaciones estadísticas del Instituto Nacional de Estadística desde el año 2010, y que desgraciadamente aún no dispone de datos publicados. Prueba de ello es que no existe una única fuente. Para conocer y analizar el gasto farmacéutico hospitalario es necesario acudir a dos fuentes diferentes: IMS Health y Farmaindustria.
- l. El gasto farmacéutico por recetas ha experimentado un descenso medio anual acumulado del 4,02% en el periodo 2008-2014, resultado de las medidas de contención del gasto implantadas, mientras que el número de recetas se ha mantenido en el periodo, apenas un descenso del 0,40%, por lo que se deduce que lo que ha bajado es el precio de los medicamentos entre 2008 y 2014.
- m. Del análisis del gasto farmacéutico en recetas a nivel de Comunidad Autónoma, se concluye que el comportamiento ha sido homogéneo en el periodo 2008-2014. La cuota del gasto por Comunidad se mantiene constante en el periodo estudiado. Obteniéndose una relación lineal muy alta de tipo directo entre la cuota del gasto de la Comunidad sobre el total nacional y la cuota poblacional de ésta. Comparando el porcentaje poblacional en el 2014 con la cuota media del gasto en el periodo y con la cuota del gasto en el 2014, se deduce un comportamiento homogéneo, salvo en el caso de Madrid que tiene una

participación sobre el total de facturas inferior a su porcentaje poblacional, y Galicia y Andalucía que la tienen superior.

- n. El mercado del autocuidado (*consumer health*) en España en el periodo 2010-2014, tiene un comportamiento a la baja pero de forma muy dispar según sus segmentos. El primer repunte, tanto en valores como en unidades, se da en el 2013.
- o. Con respecto a la rentabilidad del negocio de la oficina de farmacia española, en el mercado regulado de prescripción, a día de hoy el de mayor cuota de facturación en la farmacia española, como ha quedado demostrado, se han desplegado 29 leyes desde 1996 hasta 2012, que han moderado, en menor o mayor medida, la tasa de crecimiento de la facturación de las oficinas de farmacia. Produciéndose un descenso del margen neto de las farmacias hasta llegar a un 6,9 % en el 2012 (considerando farmacias de tipo medio en España).
- p. De la valoración de la farmacia por los españoles, resultante del análisis de cuatro encuestas publicadas sobre la percepción de la oficina de farmacia en España, caben destacar los siguientes aspectos comunes:
 - Valoración: la valoración general de los usuarios supera el 8 sobre 10 en todas las ocasiones.
 - Proximidad: más del 85% de las personas encuestadas declara tener una farmacia de referencia que es la más cercana a su domicilio; la llamada “farmacia de cabecera”.
 - Centro sanitario de referencia: la oficina de farmacia es identificada como un centro sanitario y en más del 10% de las ocasiones es visitado en primer lugar para resolver dolencias leves.
 - El servicio y los profesionales: son altamente valorados, principalmente por el trato recibido, la profesionalidad, la capacidad de resolución de dudas sanitarias, el asesoramiento y consejo y la disponibilidad de medicamentos.
- q. De la encuesta realizada, en esta investigación, respecto al interés del colectivo farmacéutico en la provincia de Sevilla en diferenciar los términos “paciente” y

“cliente” y promover relaciones con ellos, podemos concluir que el sector en la provincia de Sevilla está sensibilizado con este tema. La propia situación del sector y la promoción de la venta de medicamentos sin receta por internet en el verano de 2014 por parte de la Agencia Española de Medicamentos y Productos Sanitarios son hechos bastante notorios como para estimular al conjunto de farmacéuticos comunitarios a permanecer alerta ante nuevas formas de desarrollo de su negocio.

- r. Como conclusión, en el apartado de caracterización de la oficina participante de forma voluntaria, podemos afirmar que se trata principalmente de una oficina de farmacia urbana, que no es única en su municipio y cuya distancia al centro de salud más próximo es superior a 250 metros. La mayoría de las oficinas de farmacia participantes, el 59,02%, distinguen los términos “cliente” y “paciente”. Se trata por tanto de un farmacéutico que se enfrenta a una competencia notable y está suficientemente sensibilizado con la situación de su negocio.
- s. La característica “ser o no la única farmacia en la localidad” se ha revelado un factor discriminante para participar o no en el estudio.
- t. Las oficinas de farmacia participantes reconocen en su mayoría distinguir entre clientes y pacientes, que el medio de comunicación con ellos es el teléfono, seguido de lejos por el *SMS*. Este hecho es indicativo la fase temprana en la que se encuentra el sector, al menos en Sevilla para las farmacias participantes, respecto al uso de herramientas de comunicación con sus clientes.

Resulta lógica la perturbación que sufren los farmacéuticos porque van a tener que adaptarse a múltiples cambios, a nuevos competidores y a nuevas formas de gestión y relación con sus clientes. Por una parte, se tienen que enfrentar a la transformación de su sector a lo que hay que sumar las nuevas reglas del mercado, con un consumidor más exigente, más infiel, con menor capacidad de gasto y más informado, y la revolución de las nuevas tecnologías.

4.3 Limitaciones de la investigación

Respecto a las limitaciones de esta investigación, podemos citar las siguientes:

- La no disponibilidad de información unificada sobre clientes y pacientes en los programas de gestión.
- Las herramientas propuestas, y su aplicación, dependen de las oficinas de farmacia donde se utilicen.
- El algoritmo propuesto, tal y como se ha diseñado en esta investigación, se aplica sobre variables cuantitativas. Para aplicarlo a variables cualitativas habría que realizar transformaciones sobre ellas.
- La matriz DAFO construida en esta investigación es la resultante de la información consultada y expuesta en este trabajo, con el conocimiento previo y experiencia del autor de la misma.
- La encuesta realizada a los farmacéuticos de Sevilla no ha sido propuesta para estimar parámetros poblacionales, solo lo ha sido para “pulsar” la opinión del colectivo en la provincia de Sevilla.

4.4 Trabajos futuros y nuevas líneas de investigación

Como nuevas líneas propuestas a la finalización de esta investigación, se formulan las siguientes:

- Ampliación del algoritmo propuesto a variables de tipo cualitativo.
- Investigar nuevas variaciones para la generación de poblaciones iniciales.
- Incorporar a DECCS un esquema de inclusión de preferencias del usuario para indicar si desea un número fijo de grupos.
- Extensión de DECCS a un enfoque multi-objetivo con el que determinar un conjunto de soluciones eficientes.
- Indagar y proponer nuevas operaciones de cruce y mutación que provoquen una mayor diversidad en la búsqueda de soluciones.
- Una vez se produzca la consideración del término “cliente” en los programas de gestión de las oficinas de farmacia, se debería profundizar en las variables discriminantes en el comportamiento del consumo de clientes y pacientes.
- Abordaje de nuevas técnicas estadísticas aplicadas a las oficinas de farmacia sobre la base de la segmentación propuesta en esta investigación.
- Ahondar en la aplicación de técnicas de fidelización en el entorno de la oficina de farmacia y desarrollo de sistemas para la gestión de la relación con sus clientes.
- Ampliación de la segmentación con variables socio-demográficas propias del entorno local de la oficina de farmacia.

5 BIBLIOGRAFÍA

1. Iñesta A., Oteo L.A. La industria farmacéutica y la sostenibilidad de los sistemas de salud en países desarrollados y América Latina. *Ciência & Saúde Coletiva*. 2011 Junio; 16(6).
2. Consejo General de Colegios Oficiales de Farmacéuticos. Los modelos de Farmacia en la Unión Europea. Madrid; 2007.
3. Serra J.C. Todo lo que debe saber sobre marketing farmacéutico: Profit; 2010.
4. Salvador J.C. Gestión sanitaria integral: pública y privada: Centro de Estudios Financieros; 2010.
5. Dixon A. Facing the challenges of health care financing in the CEE/NIS. En Draft Paper-US AID Conference; 2002; Washington D. C.
6. Elola J. Health Care System Reforms in Western European Countries: The Relevance of Health Care Organization. *International Journal of Health Services*. 1996; 26(2): 239-251.
7. PricewaterhouseCoopers. HealthCast "La personalización del diagnóstico, la asistencia y el tratamiento". Informe sectorial. Health Research Institute de PricewaterhouseCoopers; 1999.
8. Boletín Oficial del Estado. Real Decreto-ley 16/2012, de 20 de abril, de medidas urgentes para garantizar la sostenibilidad del Sistema Nacional de Salud y mejorar la calidad y seguridad de sus prestaciones. 2012 Abril 20. BOE núm 98 de 24/04/2012.
9. Antoñanzas Villar F., Rodríguez-Ibeas R., Juárez-Castelló C.A., Lorente Antoñanzas M.R. Impacto del real decreto-ley 16/2012 sobre el copago farmacéutico en el número de recetas y en el gasto farmacéutico. *Revista Española de Salud Pública*. 2014; 88(2).
10. Riu I. La farmacia en Estados Unidos. OFFARM. 2005 Marzo; 24(3).

11. Fernández A.L. Productos farmacéuticos en Japón. Estudio de mercado. Madrid: Comunidad de Madrid, Consejería de Economía e Innovación Tecnológica; 2007.
12. Aula de opinión sanitaria. Comité de expertos. Ordenación Farmacéutica. Miércoles 16 de febrero de 2005.
13. Cabezas López M.D., Piqueras García A., Gómez Beteta J.E. El modelo Mediterráneo de Farmacia y la libertad de establecimiento en la Unión Europea. *Ars Pharmaceutica*. 2010 Junio;(51): 1-13.
14. Granda E. Modelos de farmacia: comparativa europea. *Economía en salud*. 2011 Marzo-Abril; 25(2).
15. Consejo General de Colegios Oficiales de Farmacéuticos. www.portalfarma.com. [Online]; 2015 [citado 24 3 2015. Disponible en: <http://www.portalfarma.com/Profesionales/organizacionfcolegial/profesionfarma/Paginas/colegiaciondatestadisticos.aspx>.
16. Muy Ilustre Colegio Oficial de Farmacéuticos de Valencia. Modelo Mediterráneo. *Cuadernos de Farmacia*. 2008 Junio;(201).
17. Consejo General de Colegios Oficiales de Farmacéuticos. Portalfarma. [Online]; 2015 [citado 23 3 2015. Disponible en: <http://www.portalfarma.com/profesionales/infoestadistica/Paginas/listado.aspx>.
18. Consejo General de Colegios Oficiales de Farmacéuticos. www.portalfarma.com. [Online]; 2013 [citado 23 5 2015. Disponible en: http://www.portalfarma.com/Profesionales/infoestadistica/Documents/Medicamento_Farmacia_Cifras_2013.pdf.
19. Muy Ilustre Colegio Oficial de Farmacéuticos de Valencia. Una imagen vale más que mil palabras. *Cuadernos de Farmacia*. 2008 Mayo;(200).

20. Boletín Oficial del Estado. Ley 29/2006, de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios. 2006. «BOE» núm. 178, de 27 de julio de 2006.
21. Boletín Oficial del Estado. Ley 16/1997, de 25 de abril, de Regulación de Servicios de las Oficinas de Farmacia. 1997. «BOE» núm. 100, de 26 de abril de 1997, páginas 13450 a 13452 (3 págs.).
22. Asociación Española de Derecho Sanitario. El modelo de farmacia español frente a otras propuestas. En Congreso de Derecho Sanitario; 2008; Madrid.
23. Meneu R. La distribución y dispensación de medicamentos. Gaceta Sanitaria. 2006; 20(Sup. 1): 154-159.
24. BOE núm. 131, 30 mayo 2008. Real Decreto 823/2008, de 16 de mayo, por el que se establecen los márgenes, deducciones y descuentos correspondientes a la distribución y dispensación de medicamentos de uso humano. 2008 May 16..
25. FEDIFAR – Federación de Distribuidores Farmacéuticos. Análisis sectorial de la distribución farmacéutica en España. Madrid; 2013.
26. Rovira Forns J., Gómez Pajuelo P., del Llano Señarís J. La regulación del precio de los medicamentos en base al valor Casal FG, editor. Madrid: Fundación Gaspar Casal; 2012.
27. Hudson S. Hospital pharmacies. En Mossialos E., Mrazek M., Walley T., editors. Regulating pharmaceuticals in Europe: striving for efficiency, equity and quality.: Open University Press; 2004.
28. Catalá Pizarro R.M, González del Tánago del Río S. Dispensación de medicamentos. 2002.
29. Gil Navarro M.V. Comunicación al 2º Curso de Evaluación y Selección de Medicamentos. 2005; Sevilla.

30. Fundamed. Experiencia española. Efectos de los cambios retributivos. Madrid; 2005.
31. Consejo General de Colegios Oficiales de Farmacéuticos. www.portalfarma.com. [Online]; 2015 [citado 28 3 2015 [(Última modificación: 05/02/2015 14:40)]. Disponible en: <http://www.portalfarma.com/Profesionales/organizacionfcolegial/profesionfarma/Paginas/colegiaciondatestadisticos.aspx>.
32. Boletín Oficial del Estado. Real Decreto-ley 5/2000, de 23 de junio, de Medidas Urgentes de Contención del Gasto Farmacéutico público y de racionalización del Uso de los Medicamentos. 2000. BOE núm. 151 de 24 de Junio de 2000.
33. Boletín Oficial del Estado. Real Decreto-ley 8/2010, de 20 de mayo, por el que se adoptan medidas extraordinarias para la reducción del déficit público. 2010. BOE Núm. 126 Lunes 24 de mayo de 2010 Sec. I. Pág. 45070.
34. Boletín Oficial del Estado. Real Decreto-ley 9/2011, de 19 de agosto, de medidas para la mejora de la calidad y cohesión del sistema nacional de salud, de contribución a la consolidación fiscal, y de elevación del importe máximo de los avales del Estado para 2011. 2011. Núm. 200 Sábado 20 de agosto de 2011 Sec. I. Pág. 93143.
35. Lillo Fernández de Cuevas, J.M.; Rodríguez Blas, M.C. Estadística de Gasto Sanitario Público. Madrid: Ministerio de Sanidad, Servicios Sociales e Igualdad; 2012.
36. Altemir, J. Características del mercado de los medicamentos de prescripción. En Serra J C, editor. Todo lo que debe saber sobre marketing farmaceutico. Madrid: Profit; 2010:73-90.

37. Ordieres E. Características del mercado de los medicamentos genéricos. En Serra JC, editor. Todo lo que debe saber sobre marketing farmacéutico. Madrid: Profit; 2010:91-102.
38. San Miguel Inza F. El mercado del medicamento en España: exploración del sector. Colección "Informes y Estudios". Institución Futuro; 2005. Informe 7.
39. Tobar F. Referenciación internacional sobre políticas de regulación de precios de medicamentos. Informe final del Convenio Interadministrativo 292/197052. Ministerio de Protección Social de la República de Colombia; 2011.
40. Maynard A, Bloor K. Dilemas in the regulation of the market for pharmaceuticals. Health Affairs. 2003 Mayo-Junio; 22(3).
41. Gómez Martínez M.E., Ruiz Romero J.A., Martínez Olmo J. Políticas de uso racional del medicamento en Europa. Revista de Administración Sanitaria. 1999 Enero-Marzo; 3(9).
42. Zara Yahni C., Segú Tolsa J.L., Font Pous M., Rovira Forns J. La regulación de los medicamentos: teoría y práctica. Gaceta Sanitaria. 1998 Enero-Febrero; 12(1).
43. Atmetlla E. Estructura competitiva de la industria farmacéutica de los medicamentos de prescripción. En Serra JC, editor. Todo lo que debe saber sobre marketing farmacéutico. Madrid: Profit; 2010: 63-71.
44. Danzon P.M. Competencia en la Industria Farmacéutica. Estudio de sector. Centro de Investigación y Docencia Económica, The Wharton School (Universidad de Pennsylvania); 2014.
45. Asociación Española de Medicamentos Genéricos. <http://www.aeseg.es/>. [Online]; 2014 [citado 09 05 2015 Disponible en: <http://www.aeseg.es/es/medicamentos-genericos-en-union-europea>.
46. Cuerda R. A. El valor del medicamento genérico. 2015 Abril.

47. Farmaindustria. Análisis de la evolución del gasto farmacéutico público en España, mes a mes; 2015.
48. Chacón I, Espoz F. <http://www.hmr.pt/>. [Online]; 2014 [citado 9 5 2015]. Disponible en: <http://www.hmr.pt/es/espanol-el-mercado-farmaceutico-de-prescripcion-crece-un-24-en-los-10-primeros-meses-del-ano/>.
49. Espinosa A., Fallon M. Características del mercado de los medicamentos hospitalarios. En Serra J.C., editor. Todo lo que debe saber sobre marketing farmacéutico. Madrid: Profit; 2010: 103-130.
50. Ministerio de Sanidad SSeI. Ministerio de Sanidad, Servicios Sociales e Igualdad. [Online]. [citado 17 04 2015]. Disponible en: <http://www.msssi.gob.es/profesionales/CarteraDeServicios/ContenidoCS/5PrestacionFarmaceutica/PF-PrestacionFarmaceutica.htm>.
51. Ministerio de Sanidad SSeI. www.msssi.gob.es. [Online]. [citado 19 04 2015]. Disponible en: <http://www.msssi.gob.es/profesionales/CarteraDeServicios/ContenidoCS/3AtencionEspecializada/home.htm>.
52. Iñesta García A. www.uned.es. [Online]. Madrid; 2012 [citado 17 04 2015]. Disponible en: <http://e-spacio.uned.es/fez/eserv/bibliuned:500626/n5-9 Politicas del medicamento.pdf>.
53. Instituto Nacional de Estadística. www.ine.es. [Online]; 2010 [citado 22 04 2015]. Disponible en: http://www.ine.es/ss/Satellite?L=es_ES&c=TFichaIOE_C&cid=1259931154471&p=1254735038414&pagename=IOEhist%2FIOEhistLayout.
54. Iñesta García A. Sobre medicamentos y farmaeconomía. Manual docente. Madrid: Instituto de Salud Carlos III, Escuela Nacional de Sanidad; 2011.

55. IMS Health. [Online]. [citado 30 6 2015. Disponible en: <http://www.imshealth.com/>.
56. Farmaindustria. [Online]. [citado 30 6 2015. Disponible en: <http://www.farmaindustria.es/>.
57. Consejo Económico y Social. Desarrollo autonómico, competitividad y cohesión social en el sistema sanitario. Madrid; 2010. Informe 01/2010.
58. Ministerio de Sanidad SSeI. <http://www.msssi.gob.es/>. [Online]; 2015 [citado 17 04 2015. Disponible en: <http://www.msssi.gob.es/profesionales/CarteraDeServicios/ContenidoCS/5PrestacionFarmaceutica/PF-PrestacionFarmaceutica.htm>.
59. Martínez Jorge M. El mercado farmacéutico español. Informe sectorial. IMS Health; 2012.
60. Boletín Oficial del Estado. Real Decreto-ley 9/2011, de 19 de agosto, de medidas para la mejora de la calidad y cohesión del sistema nacional de salud, de contribución a la consolidación fiscal, y de elevación del importe máximo de los avales del Estado para 2011. 2011. «BOE» núm. 200, de 20 de agosto de 2011.
61. Boletín Oficial del Estado. Real Decreto-ley 16/2012, de 20 de abril, de medidas urgentes para garantizar la sostenibilidad del Sistema Nacional de Salud y mejorar la calidad y seguridad de sus prestaciones. 2012. «BOE» núm. 98, de 24 de abril de 2012.
62. Whitaker D., Sánchez P.L. Diagnóstico y perspectivas del gasto farmacéutico en España. Madrid; 2001.
63. OCDE. OECD Health Data: Pharmaceutical market. [Online]; 2013 [citado 03 05 2015. Disponible en: http://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics/oecd-health-data-pharmaceutical-market_data-00545-en.

-
64. ANEFP. Asociación para el Autocuidado de la Salud. <http://www.anefp.org/>. [Online]; 2015 [citado 16 05 2015. Disponible en: <http://www.anefp.org/que-es-anefp>.
65. Almarza C. Tendencias del mercado farmacéutico español. Informe de mercado. Madrid: IMS Health; 2015.
66. Navarro J. Características del mercado de los medicamentos publicitarios. En Serra JC, editor. Todo lo que debe saber sobre marketing farmacéutico. Madrid: Profit; 2010: 131-142.
67. Alfonso Gala T. El rol del farmacéutico en automedicación. Uso de medicamentos: análisis desde la experiencia en España. Pharmaceutical Care España. 1999; 1: 157-164.
68. Consejo General de Colegios Oficiales de Farmacéuticos. Panorámica de los medicamentos publicitarios en España. Farmacéuticos. 2007 Septiembre;(327).
69. ASPIME. Informe Anual de Oficinas de Farmacia 2014. Barcelona: Club de la farmacia; 2014.
70. Antares Consulting. Impacto de las medidas de recorte sobre la Oficina de Farmacia en 2013. Informe sectorial. Madrid; 2014.
71. Antares Consulting. Situación y cifras claves de la Farmacia en 2014. Informe sectorial. Madrid: Grupo Cofares; 2015.
72. Consejo General de Colegios Oficiales de Médicos de España. Informe sobre el sector farmacéutico. Informe sectorial. Madrid: Consejo General de Colegios Oficiales de Médicos de España; 2014.
73. Mensor. La realidad económica de la Farmacia en España: Estudio del impacto económico en las oficinas de farmacia de las medidas de contención del gasto

- en medicamentos en los últimos diez años. Informe sectorial. Madrid: Consejo General de Colegios Oficiales de Farmacéuticos; 2010.
74. Unión Europea. Directiva 85/432/CEE del Consejo. 16 septiembre 1985.
75. Lex Nova. Lex Nova portal jurídico. [Online]; 2010 [citado 8 6 2015. Disponible en: <http://portaljuridico.lexnova.es/jurisprudencia/JURIDICO/25812/sentencia-tjue-de-1-de-junio-de-2010-oficinas-de-farmacia-libertad-de-establecimiento>.
76. Pérez Gálvez J.F. Libertad de establecimiento de farmacias en la Unión Europea. Revista de Derecho de la Pontificia Universidad Católica de Valparaíso. 2010: 361-406.
77. Redacción Médica. Redacción Médica. [Online]; 2015 [citado 2015 junio 08. Disponible en: <http://www.redaccionmedica.com/secciones/farmacia/las-farmacias-elaboran-un-informe-para-defenderse-de-la-amenaza-de-competencia-78459>.
78. Redacción Médica. Redacción Médica. [Online]; 2015 [citado 2015 junio 08. Disponible en: <http://www.redaccionmedica.com/secciones/farmacia/la-cnmc-investiga-las-restricciones-autonomicas-en-el-mercado-de-las-oficinas-de-farmacia-76785>.
79. Cámara de comercio de Zaragoza. Estudio de Satisfacción del Usuario de la Oficina de Farmacia en Zaragoza. Estudio estadístico. Zaragoza; 2008.
80. García de la Cruz Herrero JJ. Percepción y valoración social de la Farmacia. Informe sociológico. Madrid: Universidad Autónoma de Madrid; 2010.
81. Colegio Oficial de Farmacéuticos de Sevilla. Portal Web del Colegio Oficial de Farmacéuticos de Sevilla. [Online]; 2014 [citado 14 6 2015. Disponible en: http://www.farmaceuticosdesevilla.es/estudio-sobre-la-percepcion-social-de-la-farmacia-en-sevilla_aa329.html.

82. Hill CWL, Jones G.R. Administración estratégica: un enfoque integrado. 6th ed. México: McGraw-Hill Interamericana; 2005.
83. Moreno Romero E. Portal de la Junta de Andalucía. [Online]; 2007 [citado 20 6 2015. Disponible en: http://www.juntadeandalucia.es/averroes/convivenciaeigualdad/guia/elementos/a_diagnostico/dafo.pdf.
84. IMS Health. Tendencias del Mercado Español. Informe sectorial. Madrid: IMS Health; 2014.
85. Ideas. <http://www.ideasestrategicas.es/>. [Online]; 2013 [citado 20 6 2015. Disponible en: <http://www.ideasestrategicas.es/#!Marketing-en-farmacias-%C2%BFC%C3%B3mo-transformar-las-amenazas-en-oportunidades/c141y/A313C99B-3FB0-4342-80DE-D38082E50D94>.
86. Gonzalez Pérez J. El futuro de la oficina de farmacia. Derecho y Salud. 1997; 5.
87. Cortes Generales. Constitución Española. 1978. Artículo 43.
88. Regent P. Zona de confort. Revista de Antiguos Alumnos del IEEM. 2011 Febrero.
89. Inknowation. [Video].; 2014 [citado 21 6 2015. Disponible en: https://www.youtube.com/watch?v=i07qz_6Mk7g.
90. Frutos Sastre MJ. Dispensación de productos parafarmacéuticos. Segunda ed.: McGraw-Hill; 2013.
91. Consejo General de Colegios Oficiales de Farmacéuticos. Portalfarma. [Online]; 2013 [citado 20 6 2015. Disponible en: http://www.portalfarma.com/profesionales/organizacionfcolegial/profesionfarma/planfuturofarmaciacomunitaria//Documents/Servicios%20Farmaceuticos_mayo%202013.pdf.

92. TEVA. Observatorio de la cartera de servicios desde la oficina de Farmacia. [Online]; 2014 [citado 20 6 2015. Disponible en: <http://www.observatoriocarteraservicios.com/>.
93. Real Academia Española. RAE. [Online]; 2015 [citado 28 6 2015. Disponible en: <http://www.rae.es/>.
94. American Marketing Association. AMA. [Online]; 2014 [citado 28 6 2015. Disponible en: <https://www.ama.org/resources/Pages/Dictionary.aspx?dLetter=C>.
95. The Chartered Institute of Marketing. CIM. [Online]; 2015 [citado 28 6 2015. Disponible en: <http://www.cim.co.uk/>.
96. Pujol Bengoechea B. Dirección de Marketing Madrid: Cultural; 1999.
97. Borrás Vives R. Blog personal de Rafael Borrás. [Online]; 2012 [citado 28 6 2015. Disponible en: <http://rafaelborrasv.com/2012/03/04/farmacia-pacientes-o-clientes/>.
98. Godás L. Análisis de la oficina de farmacia: El mercado y su entorno. OFFARM. 2006 Enero; 25(1).
99. Gummesson E. Total Relationship Marketing. 2nd ed.: Butterworth-Heinemann; 2002.
100. Córdoba López JF. Del marketing transaccional al marketing relacional. Entramado. 2009 Enero-Junio; 5(1).
101. Reinares Lara P, Calvo Fernández S. Gestión de la comunicación comercial: McGraw-Hill; 1999.
102. Cardona J. Farmacéutico. En Serra J.C., editor. Todo lo que debe saber sobre marketing farmacéutico. Madrid: Profit; 2010. 631-636.

103. Gitomer J. Customer Satisfaction Is Worthless, Customer Loyalty Is Priceless: How to Make Customers Love You, Keep Them Coming Back and Tell Everyone They Know: Bard Press; 1998.
104. Hernández Fernández A., Martínez García J.M. Marketing sanitario Madrid: ESIC Editorial; 2014.
105. Huete L.M. Servicios & Beneficios Barcelona: Deusto; 2003.
106. Reichheld F.F., W. Earl Sasser, Jr.. Zero Defections: Quality Comes to Services. Harvard Business Review. 1990 Septiembre-Octubre.
107. BetterBe Yourself. [Online]; 2006 [citado 6 6 2015. Disponible en: www.rrhh.magazine.com.
108. Sánchez Reyes M. Satisfacción y fidelización de los usuarios de la oficina de farmacia. 2010. Tesis doctoral.
109. rivassanti.net. [Online] [citado 10 6 2015. Disponible en: <http://www.rivassanti.net/curso-ventas/fidelizacion-de-clientes.php>.
110. De la Fuente L. Club de la farmacia. [Online]; 2012 [citado 5 3 2015. Disponible en: <https://www.clubdelafarmacia.com/blogclub/gestion/%E2%80%9Cmis-clientes-siempre-vuelven%E2%80%9D-el-plan-de-fidelidad-en-la-oficina-de-farmacia-parte-1-de-3/>.
111. Küster Boluda I. Captación y fidelización de los clientes pacientes en un entorno competitivo. En Hernández Fernández A., Martínez García J.M. Marketing sanitario. Madrid: ESIC Editorial; 2014. 59-82.
112. Sánchez Reyes M., Moreno Toral E. La fidelización de los usuarios en la oficina de farmacia. OFFARM. 2003 Mayo; 22(5).
113. Díez Cremades M.L. Atención y fidelización de clientes en la oficina de farmacia. Farmacia profesional. 2011 Septiembre-Octubre; 25(5).

114. Pharma PT. Higia. [Online]; 2013 [citado 18 12 2014. Disponible en: <http://www.higiacomunidad.es/blog/fidelizacion-en-la-oficina-de-farmacia-los-cuatro-frenos>.
115. Ojeda Casares M. Caracterización de clientes e identificación del paciente como bases de una gestión actualizada en la Oficina de Farmacia. 2013. II Jornada Científico Profesional Cátedra Avenzoar.
116. Maqueda Lafuente J. Marketing para los nuevos tiempos Sánchez C., editor.: McGraw-Hill; 2012.
117. Andrades A. Cinco días. [Online]; 2015 [citado 27 6 2015. Disponible en: http://cincodias.com/cincodias/2015/02/06/empresas/1423227182_693563.html.
118. Chamberlain EH. Teoría de la competencia monopólica. 2nd ed. Económica FdC, editor; 1956.
119. Robinson J. La economía de la competencia imperfecta. Aguilar; 1933.
120. Smith W.R. Product Differentiation and Market Segmentation as Alternative Marketing Strategies. Journal of Marketing. 1956 Julio; 26(1): 3-8.
121. Rábade L.A, Salvatierra S. Segmentación de mercados en marketing. En Sánchez C., editor. Marketing para los nuevos tiempos. Madrid: McGraw-Hill/Interamericana de España, S.L.; 2012. 27-42.
122. Oracle. Gaining Competitive Advantage Through Effective Retail Loyalty Programs. An Oracle White Paper; 2006.
123. Valderrey Sanz P. Técnicas de segmentación de mercados: Starbook Editorial; 2010.
124. Ferrell O.C., Hartline M.D. Estrategia de marketing. Quinta edición: Cengage Learning; 2012.

125. Godás L. Segmentación del mercado de la oficina de farmacia. OFFARM. 2006 Febrero; 25(2).
126. Kotler P. Dirección de marketing: análisis, planificación, gestión y control. Séptima edición: Prentice-Hall; 1995.
127. Paniagua S. Sorayapaniagua. [Online]; 2012 [citado 9 6 2015. Disponible en: <http://www.sorayapaniagua.com/2012/03/27/la-gran-apuesta-de-coca-cola-por-los-modelos-predictivos/>.
128. Silver N. FiveThirtyEight. [Online]; 2015 [citado 9 6 2015. Disponible en: <http://fivethirtyeight.com/contributors/nate-silver/>.
129. Paul R, Hoque ASML. Clustering medical data to predict the likelihood of diseases. Digital Information Management (ICDIM), 2010 En Fifth International Conference. 2010 Julio.
130. EFE. 20minutos. [Online]; 2013 [citado 9 6 2015. Disponible en: <http://www.20minutos.es/noticia/1979004/0/big-data/datos-masivos/delata-defraudadores/>.
131. Pascual MG. CincoDías. [Online]; 2015 [citado 9 6 2015. Disponible en: http://cincodias.com/cincodias/2015/05/01/sentidos/1430493583_087196.html.
132. Xu R,, Wunsch I,D. Survey of clustering algorithms. IEEE Transactions on Neural Networks. 2005; 16(3).
133. Berkhin P. A. Survey of Clustering Data Mining Techniques. En Kogan J., Nicholas C, Teboulle M., editors. GROUPING MULTIDIMENSIONAL DATA: Springer Berlin Heidelberg; 2006. 25-71.
134. Das S., Abraham A., Konar A. Metaheuristic Clustering: Springer-Verlag Berlin Heidelberg; 2009.

135. Dimitriadou E, Dolnicar S, Weingassel A. An Examination of Indexes for Determining the Number of Clusters in Binary Data Sets. *Psychometrika*. 2002; 67(1): 137-160.
136. Vendramin L., Campello R.J.G.B., Hruschka E. Relative Clustering Validity Criteria: A Comparative Overview. Wiley InterScience. 2010 Mayo.
137. Tibshirani R., Walther G., Hastie T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2001; 63(2): 411–423.
138. Desgraupes B. <http://cran.r-project.org/>. [Online]; 2013. Disponible en: <http://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf>.
139. Calinski R.B., Harabasz J. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*. 1974; 3: 1-27.
140. Kolesnikov A., Trichina E., Kauranne T. Estimating the number of clusters in a numerical data set via quantization error modeling. *Pattern Recognition*. 2015 Marzo; 48(3).
141. Subbalakshmi C., Krishna G.R., Rao SKM, Rao P.V. A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set. *Procedia Computer Science*. 2015; 46: 346–353.
142. Kryszczuk K., Hurley P. Estimation of the Number of Clusters Using Multiple Clustering Validity Indices. *Lecture Notes in Computer Science*. 2010; 5997: 114-123.
143. Von Luxburg U. A Tutorial on Spectral Clustering. *Statistics and Computing*. 2007; 17(4).
144. Handl J., Knowles J. Multiobjective clustering with automatic determination of the number of clusters. 2004. Informe técnico.

145. Matake N., Hiroyasu T., Miki M., Senda T. Multiobjective clustering with automatic k-determination for large-scale data. En GECCO '07 Proceedings of the 9th annual conference on Genetic and evolutionary computation; 2007; London. 861-868.
146. Suresh K., K.D., Ghosh S., Das S., Abraham A., Han S.Y. Multi-objective differential evolution for automatic clustering with application to micro-array data analysis. *Sensors*. 2009; 9(5): 3981-4004.
147. Davies D.L., Bouldin D.W. A Cluster Separation Measure. En *IEEE Transactions Pattern Analysis and Machine Intelligence*. 1979 Mayo: 224-227.
148. Dunn J.C. Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*. 1974 Abril; 4(1): 95-104.
149. Chou C.H., Su M.C., Lai E. A new cluster validity measure and its application to image compression. *Pattern Analysis and Applications*. 2004 Julio; 7(2): 205-220.
150. Handl J., Knowles J. An Evolutionary Approach to Multiobjective Clustering. *Evolutionary Computation*, En *IEEE Transactions*. 2007 Febrero; 11(1): 56-76.
151. Saha S., Bandyopadhyay S. A symmetry based multiobjective clustering technique for automatic evolution of clusters. *Pattern Recognition*. 2010; 43(3): 738-751.
152. Law M.H.C., Topchy A.P., Jain A.K. Multiobjective data clustering. *Computer Vision and Pattern Recognition*, 2004. CVPR 2004. En *Proceedings of the 2004 IEEE Computer Society Conference*. 2004 Julio; 2: 424-430.
153. Handl J., Knowles J. Multiobjective clustering around medoids. En *Evolutionary Computation*, *IEEE Transactions*. 2005 Septiembre; 1: 632-639.

154. Saha S., Bandyopadhyay S. Automatic MR brain image segmentation using a multiseed based multiobjective clustering approach. *Applied Intelligence*. 2010 Diciembre; 35(3): 411-427.
155. Kaufman L., Rousseeuw P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*: Wiley; 1990.
156. Sosa García J., Vega Pons S., Ruiz Shulcloper J. Algoritmos de agrupamiento difuso, índices de validación: un estado del arte. Reporte técnico. La Habana: CENATAV, Dpto. Minería de Datos y reconocimiento de patrones; 2012. Report No.: ISSN 2072-6287.
157. Bezdek J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms* New York: Plenum Press; 1981.
158. Lloyd S.P. Least Squares Quantization in PCM. *En IEEE Transactions Information Theory*. 1982 Marzo; 28: 129-137.
159. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. 1967; 1: 281-297.
160. Izenman A.J. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. New York: Springer-Verlag New York; 2008.
161. Han J., Kamber M., Tung A.K.H. Spatial Clustering Methods in Data Mining: A Survey. *En Geographic Data Mining and Knowledge Discovery: Taylor and Francis*; 2001.
162. Ng RT, Han J. Efficient and Effective Clustering Methods for Spatial Data Mining. *En Proceedings of the 20th International Conference on Very Large Data Bases*; 1994; San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 144–155.

163. Pakhira M.K. Digital image processing and pattern recognition New Delhi: PHI Learning; 2011.
164. Jain A., Dubes R. Algorithms for Clustering Data New Jersey: Prentice-Hall; 1988.
165. Lance G., Willians W. A general theory of classification sorting strategies. Computer Journal. 1967; 9: 373-386.
166. Zhang T, Ramakrishnan R, Livny M. Birch: an efficient data clustering method. En SIGMOD '96: Proceedings of the 1996 ACM SIGMOD international; 1996; Montreal. 103-114.
167. Karypis G., Han E.H, Kumar V. Chameleon: A hierarchical clustering algorithm using dynamic modeling. 68-75.
168. Glover F. Future Paths for Integer Programming and Links to Artificial Intelligence. Computers & Operations Research. 1986; 13(5): 533-549.
169. Alba E. Parallel Metaheuristics: A New Class of Algorithms: Wiley; 2005.
170. Blum C., Roli A. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. ACM Computing Surveys. 2003 Septiembre; 35(3): 268-308.
171. Siarry P., Michalewicz Z., editors. Advances in Metaheuristics for Hard Optimization: Springer-Verlag Berlin Heidelberg; 2008.
172. Kirkpatrick S, Gelatt C.D., Vecchi M.P. Optimization by Simulated Annealing. Science, New Series. 1983 Mayo; 220(4598): 671-680.
173. Černý V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. Journal of Optimization Theory and Applications. 1985; 45(1): 41-51.

174. Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller E. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*. 1953; 21: 1087-1092.
175. Glover F, Laguna M. Tabu Search. En Reeves CR, editor. *Modern heuristic techniques for combinatorial problems*. New York, NY, USA: John Wiley & Sons, Inc; 1993. 70-150.
176. Feo T.A, Resende M.G.C. Greedy randomized adaptive search procedures. *Journal of Global Optimization*. 1995; 6: 109-133.
177. Resende M.G.C., Ribeiro C.C. Greedy randomized adaptive search procedures. En Glover F., Kochenberger G., editors. *Handbook of Metaheuristics.*: Kluwer Academic Publishers; 2003. 219-249.
178. Lourenço H.R., Martin O., Stützle T. Iterated Local Search. En Glover F., Kochenberger G., editors. *Handbook of Metaheuristics.*: Kluwer Academic Publishers; 2003. 321-353.
179. Hansen P., Mladenović N. An introduction to variable neighborhood search. En *Meta-heuristics.*: Springer US; 1999. 433-458.
180. Darwin C. *The Origin of Species by Means of Natural Selection or The Preservation of Favored Races in the Struggle for Life*; 1859.
181. Fogel L.J., Owens A, Walsh M. *Artificial Intelligence Through Simulated Evolution* New York: John Wiley & Sons, Inc.; 1966.
182. Holland J.H. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*: Ann Arbor : University of Michigan Press; 1975.
183. Schwefe H.P. *Numerische Optimierung von Computer-Modellen*. 1974. Tesis.

184. Rechenberg I. Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. 1971. Tesis.
185. Fogel D.B. Evolutionary Computation: Toward a New Philosophy of Machine Intelligence Piscataway, NJ, USA: IEEE Press; 1995.
186. Goldberg D.E. Genetic Algorithms in Search, Optimization and Machine Learning. Boston, MA, USA: Addison-Wesley Longman Publishing Co.; 1989.
187. Michalewicz Z., Fogel D.B. How to Solve It: Modern Heuristics. 2nd ed.: Springer-Verlag Berlin Heidelberg; 2004.
188. Fogel. L.J. Artificial intelligence through simulated evolution: John Wiley; 1966.
189. Hernández A., Buckles B.P., Coello C.A. Estrategias evolutivas: La versión alemana del algoritmo genético (Parte I). Soluciones avanzadas. 1998 Octubre; 6(62): 38-45.
190. Koza J.R. Genetic Programming: On the Programming of Computers by Means of Natural Selection Cambridge, MA, USA: MIT Press; 1992.
191. Barricelli N.A. Numerical testing of evolution theories. Acta Biotheoretica. 1962; 16(1-2): 69-98.
192. Koza J.R., Keane M.A., Streeter M.J. What's AI done for me lately? Genetic programming's human-competitive results. Intelligent Systems, IEEE. 2003 Mayo; 18(3): 25-31.
193. Storn R., Price K. Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. Journal of Global Optimization. 1997 Diciembre; 11(4): 341-359.
194. Kennedy J., Eberhart R. Particle Swarm Optimization. En Neural Networks, 1995. Proceedings, IEEE International Conference; 1995; Perth, Australia. 1942-1948.

195. Kennedy J., Eberhart R. A Discrete Binary Version of the Particle Swarm Algorithm. En Proceedings of the IEEE International Conference on Systems, Man and Cybernetics; 1997. 4104-4109.
196. Kennedy J. The Particle Swarm: Social Adaptation of Knowledge. En IEEE International Conference on Evolutionary Computation; 1997. 303-308.
197. Olariu S., Zomaya A.Y. Handbook Of Bioinspired Algorithms And Applications: Chapman and Hall/CRC; 2005.
198. Polya G. How to Solve It. The New Aspect of Mathematical Method: Princeton University Press; 1971.
199. Kennedy J., Eberhart R, Y. S. Swarm Intelligence San Francisco: Morgan Kaufmann Publishers; 2001.
200. Dorigo M. Optimization, Learning and Natural Algorithms. 1992. Tesis.
201. Dorigo M., Stützle T. The Ant Colony Optimization metaheuristic: algorithms applications, and advances. En Handbook of Metaheuristics: Kluwer Academics; 2003: 251-285.
202. Segura C., Coello C.A.C., Hernández-Díaz A.G. Improving the vector generation strategy of Differential Evolution for large-scale optimization. Information Sciences. 2015 Diciembre; 323(1): 106-129.
203. Santana-Quintero L.V., G Hernández-Díaz A., Molina J., Coello Coello C.A., Caballero R. DEMORS: A hybrid multi-objective optimization algorithm using differential evolution and rough set theory for constrained problems. Computers & Operations Research. 2010 Marzo; 37(3).
204. Poles S., Fu Y., Rigoni E. The Effect of Initial Population Sampling on the Convergence of Multi-Objective Genetic Algorithms. En Multiobjective

- Programming and Goal Programming.: Springer Berlin Heidelberg; 2009: 123-133.
205. Poles S., Fu Y., Rigoni E. Multi-objective design optimization using cascade evolutionary computations. *Computer Methods in Applied Mechanics and Engineering*. 2005 Agosto; 194(30-33): 3496-3515.
206. McKay M.D., Beckman R.J., Conover W.J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*. 1979 Mayo; 21(2): 239-245.
207. Swiler L.P., Wyss G.D.; 1998 [citado 31 7 2015. Disponible en: <http://prod.sandia.gov/techlib/access-control.cgi/1998/980210.pdf>.
208. UCI Machine Learning Repository. [Center for Machine Learning and Intelligent Systems]. Disponible en: <https://archive.ics.uci.edu/ml/datasets.html?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table>.
209. KEEL (Knowledge Extraction based on Evolutionary Learning). [Data sets].; 2004-2015 [citado 9 6 2015. Disponible en: <http://sci2s.ugr.es/keel/datasets.php>.
210. University of Eastern Finland. Clustering datasets [Data sets]. Disponible en: <http://cs.joensuu.fi/sipu/datasets/>.
211. Universidad de Colonia. Data Sets for Clustering Techniques [Data sets].; 2008 [citado 9 6 2015. Disponible en: <http://www.uni-koeln.de/themen/statistik/data/cluster/>.
212. Krizhevsky A., Nair V., Hinton G. [Data sets]. [citado 9 6 2015. Disponible en: <http://www.cs.toronto.edu/~kriz/cifar.html>.

- 213. Tan KC, Lee TH, Khor EF. Evolutionary algorithms with dynamic population size and local exploration for multiobjective optimization. En *Evolutionary Computation*, IEEE Transactions. 2001 Diciembre; 5(6): 565-588.
- 214. Chen T., Tang K., Chen G., Yao X. A large population size can be unhelpful in evolutionary algorithms. *Theoretical Computer Science*. 2012 Junio; 436: 54-70.
- 215. Gibbs M.S., Maier H.R., Dandy G.C. Using characteristics of the optimisation problem to determine the Genetic Algorithm population size when the number of evaluations is limited. *Environmental Modelling & Software*. 2015 Julio; 69: 226-239.
- 216. Tizhoosh H.R. Opposition-Based Learning: A New Scheme for Machine Intelligence. En *International Conference on Computational Intelligence for Modelling Control and Automation (CIMCA'2005)*; 2005; Viena. 695-701.
- 217. Rahnamayan S., Tizhoosh H.R., Salama M.M.A. Opposition-based differential evolution for optimization of noisy problems. En *IEEE World Congress on Computational Intelligence*; 2006; Vancouver. 6756–6763.
- 218. Rahnamayan S., Tizhoosh H.R., Salama M.M.A. Opposition-based differential evolution. En Chakraborty UK, editor. *Advances in Differential Evolution.*: Springer Berlin Heidelberg; 2008: 155-171.
- 219. Rahnamayan S., Tizhoosh H.R. Differential evolution via exploiting opposite populations. En Ventresca M., Tizhoosh H.R., editors. *Oppositional Concepts in Computational Intelligence*: Springer Berlin Heidelberg; 2008: 143-160.
- 220. Rahnamayan S., Tizhoosh H.R., Salama M.M.A. Opposition-Based Differential Evolution. *Evolutionary Computation*, IEEE Transactions on. 2008 Enero; 12(1): 64-79.

221. Rahnamayan S., Wang G. Investigating in Scalability of Opposition-Based Differential Evolution. En 8th WSEAS International Conference on SIMULATION, MODELLING and OPTIMIZATION (SMO '08); 2008: Santander.
222. FEDIFAR – Federación de Distribuidores Farmacéuticos. Análisis sectorial de la distribución farmacéutica en España. Madrid; 2013.
223. Farmaindustria. Analisis de la evolución del gasto farmacéutico público en España. Informe sectorial. Madrid: Farmaindustria; 2015.
224. Center for Machine Learning and Intelligent Systems. UCI Machine Learning Repository [Data sets]. Disponible en: <https://archive.ics.uci.edu/ml/datasets.html?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table>.
225. Maheswaran G., P. J., Jose J., Joseph J. K Means Clustering Algorithms : A Comparitive. www.academia.edu. 2013 Marzo.
226. Guha S., Rastogi R., Shim K. CURE: An Efficient Clustering Algorithm for Large Databases. En ACM , editor. En Proceedings of the 1998 ACM SIGMOD international conference Management of data (SIGMOD '98); 1998; Seattle. 73-84.
227. Wagstaff K., Cardie C., Rogers S, Schrödl S. Constrained K-means Clustering with Background Knowledge. En Proceedings of the Eighteenth International Conference on Machine Learning; 2001. 577-584.
228. De Amorim R.C. Constrained Clustering with Minkowski Weighted K-Means. En Proceedings of the 13th IEEE International Symposium on Computational Intelligence and Informatics; 2012. 13-17.
229. Liu B., Xia Y., Yu P.S. Clustering Through Decision Tree Construction. En Proceedings of the Ninth International Conference on Information and Knowledge Management; 2000; McLean, Virginia, USA. 20-29.

- 230. Varmuza K., Filzmoser P. Introduction to multivariate statistical analysis in chemometrics: CRC Press (Taylor & Francis); 2009.
- 231. Zanakins S.H., Evans J.R. Heuristic "Optimization": Why, When, and How to Use It. Interfaces. 1981 Octubre; 11(5): 84-91.
- 232. Almarza,C. Evolución del Mercado de Genéricos. Informe sectorial. Madrid: IMS Health; 2013.
- 233. Almarza C. Tendencias del mercado farmacéutico español. Informe sectorial. Madrid: IMS Health; 2015.

6 ANEXOS

6.1 Cuestionario tarjeta fidelización

CUESTIONARIO TARJETA SALUDABLE



Fecha _____ Profesional _____

1. Datos personales

| | | | |
|------------------|--|---------------------|--|
| Nombre | | | |
| Apellidos | | | |
| NIF | | Fecha de nacimiento | |
| Dirección | | | |
| Población | | Provincia | |
| C.P. | | Teléfono fijo | |
| Teléfono móvil 1 | | Teléfono móvil 2 | |
| e-mail | | | |
| Estudios | | | |
| Profesión | | | |

2. Estado civil

☐ Casado/a ☐ Soltero/a ☐ Otro _____

3. Sexo

☐ Hombre ☐ Mujer

4. Número de hijos _____

5. Beneficiarios: si deseas que otras personas acumulen puntos en tu cuenta, anota aquí sus nombres y tu relación con ellas.

| Nombre | Relación | Edad | |
|--------|----------|------|-------|
| | | Años | Meses |
| | | | |
| | | | |
| | | | |
| | | | |

Los beneficios se acumularán en la cuenta del titular principal

6. ¿Con qué frecuencia sueles ir a ésta u otras farmacias?

- ☐ Una vez a la semana ☐ Una vez cada quince días
☐ Una vez al mes ☐ Una vez cada dos meses
☐ Una vez cada tres meses ☐ Una vez cada seis meses

7. ¿Cuáles de los siguientes productos sueles comprar cuando acudes a la farmacia?

- ☐ Medicamentos ☐ Cosmética
☐ Higiene dental ☐ Productos infantiles
☐ Dietética ☐ Higiene corporal
☐ Homeopatía ☐ Plantas medicinales
☐ Otras (indicar cuál) _____

8. ¿Eres fumador/a? ☐ Sí ☐ No

9. ¿Qué partes de tu aspecto cuidas?

- ☐ Cabello ☐ Piel ☐ Rostro ☐ Peso
☐ Dientes ☐ Manos ☐ Pies ☐ Ojos
☐ Otras (indicar cuáles) _____

10. ¿Qué haces para cuidar tu salud?

- ☐ Hago deporte habitualmente
☐ Me gusta estar informado y leer publicaciones relacionadas con temas sanitarios
☐ Realizo analíticas y chequeos periódicos
☐ Controlo mi alimentación
☐ Realizo revisiones periódicas de salud ocular
☐ Solicito información al médico cuando tengo dudas
☐ No hago nada especial para cuidar mi salud
☐ Otras (indicar cuáles) _____

11. ¿Cuáles de estos servicios te parecen interesantes?

- ☐ Seguimiento de la medicación
☐ Sistema personalizado dosificación de la medicación
☐ Asesoramiento cosmético
☐ Asesoramiento deshabituación tabáquica
☐ Asesoramiento productos infantiles
☐ Asesoramiento en homeopatía
☐ Asesoramiento en dietas y nutrición
☐ Asesoramiento en cuidados visuales
☐ Asesoramiento en medicina natural y herboristería
☐ Atención farmacéutica domiciliaria
☐ Otras (indicar cuáles) _____

12. ¿Sobre qué temas te gustaría recibir información?

- ☐ Cuidados de la vista ☐ Dietas personalizadas
☐ Fitoterapia ☐ Homeopatía
☐ Estilo de vida cardiosaludable ☐ Dermocosmética
☐ Cuidado pediátrico ☐ Ninguno
☐ Otras (indicar cuáles) _____

13. ¿Por qué has elegido esta farmacia para realizar tus compras?

- ☐ Precios adecuados ☐ Recomendada por otros
☐ Variedad de productos ☐ Proximidad/accesibilidad
☐ Atención profesional ☐ Suelo comprar en otras

☐ No quiero que se me envíe información sanitaria o promociones de mi interés de esta Oficina de Farmacia.

La información que se le solicita es facultativa, si bien, de no facilitarla no se le podrá prestar el servicio. Usted queda informado de que el anterior consentimiento es revocable y de que podrá ejercer sus derechos de acceso, rectificación, cancelación y oposición de conformidad con lo dispuesto en la legislación vigente en materia de protección de datos dirigiéndose por escrito a la siguiente dirección C/ Travesía 000, 9, 40008, 41015 Sevilla.

Firma del cliente

6.2 Muestreos realizados para 20, 30, 40 y 100 soluciones

6.2.1 Problema IRIS

| EJECUCIÓN | ALEATORIO | | | | HIPERCUBOS | | | | VECTORES OPUESTOS | | | |
|--------------|-----------|--------|--------|--------|------------|--------|--------|--------|-------------------|--------|--------|--------|
| | N= 20 | N= 30 | N= 40 | N= 100 | N= 20 | N= 30 | N= 40 | N= 100 | N= 20 | N= 30 | N= 40 | N= 100 |
| Mejor | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,7659 |
| Media | 0,8293 | 0,8065 | 0,7958 | 0,7723 | 0,7852 | 0,7784 | 0,7736 | 0,7696 | 0,8915 | 0,8517 | 0,8293 | 0,7875 |
| Peor | 1,0524 | 1,0069 | 0,8692 | 0,8107 | 0,8463 | 0,8384 | 0,8139 | 0,7811 | 1,3431 | 1,3091 | 1,0524 | 0,8424 |
| 1 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,8346 | 0,7705 | 0,7751 | 0,7659 | 0,8285 | 0,7659 | 0,7659 | 0,7659 |
| 2 | 0,7832 | 0,7659 | 0,7659 | 0,7842 | 0,7821 | 0,7713 | 0,7821 | 0,7713 | 0,7659 | 0,7832 | 0,7832 | 0,7768 |
| 3 | 0,7659 | 0,7811 | 0,7811 | 0,7659 | 0,7659 | 0,7811 | 0,7705 | 0,7737 | 0,7832 | 0,7659 | 0,7659 | 0,7842 |
| 4 | 0,8043 | 0,8109 | 0,7842 | 0,7768 | 0,7741 | 0,7775 | 0,7737 | 0,7705 | 0,8384 | 0,7768 | 0,8043 | 0,8285 |
| 5 | 0,7811 | 0,7842 | 0,8285 | 0,7659 | 0,7768 | 0,7726 | 0,7659 | 0,7659 | 0,7659 | 0,8463 | 0,7811 | 0,8285 |
| 6 | 0,8109 | 0,8285 | 0,8285 | 0,8107 | 0,7747 | 0,7768 | 0,7741 | 0,7659 | 0,7768 | 0,7811 | 0,8109 | 0,7659 |
| 7 | 0,7842 | 0,8285 | 0,7659 | 0,7751 | 0,8285 | 0,8285 | 0,7705 | 0,7713 | 0,8463 | 0,8463 | 0,7842 | 0,8384 |
| 8 | 0,8384 | 0,8285 | 0,8312 | 0,7659 | 0,7713 | 0,7832 | 0,7713 | 0,7659 | 0,8043 | 0,8109 | 0,8384 | 0,7768 |
| 9 | 0,8285 | 0,7659 | 0,8418 | 0,7768 | 0,7811 | 0,7659 | 0,7821 | 0,7751 | 0,7811 | 0,8384 | 0,8285 | 0,7705 |
| 10 | 0,8384 | 0,7821 | 0,7768 | 0,7821 | 0,7726 | 0,7858 | 0,7832 | 0,7705 | 1,2439 | 0,7842 | 0,8384 | 0,7659 |
| 11 | 0,8285 | 0,8384 | 0,7705 | 0,7821 | 0,7751 | 0,7659 | 0,7659 | 0,7713 | 0,8463 | 0,9276 | 0,8285 | 0,8107 |
| 12 | 0,8285 | 0,8418 | 0,8679 | 0,7713 | 0,7659 | 0,7705 | 0,8139 | 0,7713 | 0,8109 | 0,8285 | 0,8285 | 0,8424 |
| 13 | 0,8326 | 0,7858 | 0,7659 | 0,7659 | 0,7726 | 0,7713 | 0,7811 | 0,7659 | 0,8384 | 0,8384 | 0,8326 | 0,7751 |
| 14 | 0,7659 | 0,7768 | 0,8107 | 0,7705 | 0,7768 | 0,7659 | 0,7811 | 0,7659 | 0,7842 | 0,8285 | 0,7659 | 0,8305 |
| 15 | 0,8312 | 0,7705 | 0,8424 | 0,7705 | 0,7713 | 0,8285 | 0,7659 | 0,7811 | 0,8384 | 0,9715 | 0,8312 | 0,7821 |
| 16 | 0,8384 | 0,8679 | 0,7751 | 0,7659 | 0,8285 | 0,7821 | 0,7751 | 0,7659 | 0,9276 | 0,8285 | 0,8384 | 0,7659 |
| 17 | 0,8418 | 0,7659 | 0,8305 | 0,7659 | 0,7747 | 0,7751 | 0,7751 | 0,7713 | 0,8285 | 0,8326 | 0,8418 | 0,7768 |
| 18 | 0,8463 | 0,8107 | 0,8384 | 0,7659 | 0,7768 | 0,7821 | 0,7821 | 0,7713 | 0,9000 | 0,7659 | 0,8463 | 0,8285 |
| 19 | 0,7858 | 0,8107 | 0,7821 | 0,7811 | 0,7821 | 0,7741 | 0,7659 | 0,7768 | 0,8384 | 0,7821 | 0,7858 | 0,7821 |
| 20 | 0,7768 | 0,8424 | 0,7659 | 0,7659 | 0,7811 | 0,7659 | 0,7705 | 0,7659 | 0,8384 | 0,8312 | 0,7768 | 0,8346 |
| 21 | 0,8632 | 0,7751 | 0,7768 | 0,7705 | 0,7741 | 0,7811 | 0,7659 | 0,7713 | 0,8285 | 0,8384 | 0,8632 | 0,7821 |
| 22 | 0,7705 | 0,8285 | 0,8285 | 0,7821 | 0,8463 | 0,7741 | 0,7659 | 0,7705 | 0,9715 | 0,8463 | 0,7705 | 0,8145 |
| 23 | 0,8679 | 0,8305 | 0,8384 | 0,7659 | 0,7821 | 0,7821 | 0,7737 | 0,7659 | 1,3356 | 0,8418 | 0,8679 | 0,7751 |
| 24 | 0,8829 | 0,8463 | 0,7821 | 0,7741 | 0,7811 | 0,7737 | 0,7659 | 0,7659 | 0,8285 | 0,9659 | 0,8829 | 0,7713 |
| 25 | 0,7659 | 0,7821 | 0,8346 | 0,7659 | 0,7821 | 0,7713 | 0,7737 | 0,7659 | 0,8326 | 0,7858 | 0,7659 | 0,7741 |
| 26 | 0,8463 | 0,8285 | 0,7821 | 0,7741 | 0,7821 | 0,7659 | 0,7713 | 0,7713 | 0,8463 | 1,3091 | 0,8463 | 0,7659 |
| 27 | 0,8107 | 0,7659 | 0,8145 | 0,7713 | 0,7705 | 0,7741 | 0,7832 | 0,7726 | 0,7659 | 0,7768 | 0,8107 | 0,7705 |
| 28 | 0,8107 | 0,7789 | 0,8187 | 0,7659 | 0,8265 | 0,7821 | 0,7659 | 0,7659 | 0,7821 | 1,0047 | 0,8107 | 0,7821 |
| 29 | 0,8436 | 0,8285 | 0,7751 | 0,7713 | 0,7842 | 0,7832 | 0,7705 | 0,7705 | 0,8384 | 0,8463 | 0,8436 | 0,7741 |
| 30 | 0,8424 | 0,8384 | 0,7713 | 0,7659 | 0,8312 | 0,7659 | 0,7659 | 0,7726 | 0,8312 | 0,7705 | 0,8424 | 0,7705 |
| 31 | 1,0524 | 1,0069 | 0,7741 | 0,7821 | 0,7775 | 0,7768 | 0,7705 | 0,7751 | 0,8384 | 0,8679 | 1,0524 | 0,7659 |

| | | | | | | | | | | | | |
|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 32 | 0,7751 | 0,7821 | 0,8692 | 0,7659 | 0,7713 | 0,7821 | 0,7659 | 0,7659 | 0,8463 | 0,8829 | 0,7751 | 0,7713 |
| 33 | 0,8346 | 0,8346 | 0,7659 | 0,7751 | 0,8346 | 0,7832 | 0,7811 | 0,7737 | 1,3431 | 0,7659 | 0,8346 | 0,8285 |
| 34 | 0,8305 | 0,8312 | 0,7705 | 0,7713 | 0,7821 | 0,7713 | 0,7659 | 0,7705 | 0,8418 | 0,7821 | 0,8305 | 0,7659 |
| 35 | 0,8384 | 0,7821 | 0,8346 | 0,7726 | 0,7751 | 0,7713 | 0,7713 | 0,7751 | 0,8463 | 0,8463 | 0,8384 | 0,7842 |
| 36 | 0,8463 | 0,8145 | 0,7741 | 0,7659 | 0,8384 | 0,7713 | 0,7811 | 0,7705 | 0,9659 | 0,8107 | 0,8463 | 0,7659 |
| 37 | 0,8204 | 0,8187 | 0,8265 | 0,7713 | 0,7811 | 0,7705 | 0,7705 | 0,7659 | 0,8384 | 0,8107 | 0,8204 | 0,7821 |
| 38 | 0,7821 | 0,7751 | 0,7705 | 0,7713 | 0,7659 | 0,7705 | 0,7821 | 0,7659 | 0,7858 | 0,8463 | 0,7821 | 0,7811 |
| 39 | 0,8285 | 0,7751 | 0,7659 | 0,7659 | 0,7713 | 0,7821 | 0,7771 | 0,7659 | 1,3091 | 0,8436 | 0,8285 | 0,7659 |
| 40 | 0,7659 | 0,7713 | 0,7713 | 0,7713 | 0,7659 | 0,8285 | 0,7713 | 0,7659 | 0,7768 | 0,8424 | 0,7659 | 0,7821 |
| 41 | 0,7768 | 0,7741 | 0,8285 | 0,7775 | 0,7659 | 0,7659 | 0,7659 | 0,7659 | 0,8632 | 1,3011 | 0,7768 | 0,7726 |
| 42 | 0,7789 | 0,8692 | 0,7659 | 0,7659 | 0,7751 | 0,7659 | 0,7751 | 0,7659 | 1,1432 | 0,7751 | 0,7789 | 0,7705 |
| 43 | 0,8285 | 0,7659 | 0,7821 | 0,7811 | 0,7832 | 0,7713 | 0,7659 | 0,7659 | 0,8463 | 0,8285 | 0,8285 | 0,7821 |
| 44 | 0,8285 | 0,8312 | 0,7842 | 0,7751 | 0,7659 | 0,8384 | 0,7713 | 0,7713 | 0,7705 | 0,8346 | 0,8285 | 0,8242 |
| 45 | 0,8384 | 0,7705 | 0,7659 | 0,7713 | 0,7659 | 0,7741 | 0,7713 | 0,7737 | 1,1827 | 0,8384 | 0,8384 | 0,7659 |
| 46 | 1,0069 | 0,7821 | 0,7821 | 0,7659 | 0,7713 | 0,7659 | 0,7832 | 0,7659 | 0,8679 | 0,8305 | 1,0069 | 0,7751 |
| 47 | 1,0196 | 0,8384 | 0,8109 | 0,7659 | 0,7751 | 0,7751 | 0,7659 | 0,7705 | 0,8829 | 1,0314 | 1,0196 | 0,7741 |
| 48 | 0,7821 | 0,7741 | 0,7659 | 0,7811 | 0,7741 | 0,7659 | 0,7751 | 0,7751 | 1,2117 | 0,8463 | 0,7821 | 0,7860 |
| 49 | 0,8346 | 0,8265 | 0,7832 | 0,7659 | 0,7705 | 0,7659 | 0,7751 | 0,7713 | 0,7659 | 0,8204 | 0,8346 | 0,8285 |
| 50 | 0,9165 | 0,7705 | 0,7821 | 0,7713 | 0,7705 | 0,7821 | 0,7659 | 0,7659 | 0,7821 | 0,7821 | 0,9165 | 0,7659 |
| 51 | 0,8312 | 0,7832 | 0,7726 | 0,7741 | 0,8384 | 0,7737 | 0,7751 | 0,7705 | 1,0129 | 0,8285 | 0,8312 | 0,8139 |

6.2.2 Problema GLASS

| | ALEATORIO | | | | HIPERCUBOS | | | | VECTORES OPUESTOS | | | |
|--------------|-----------|---------|--------|--------|------------|--------|--------|--------|-------------------|--------|--------|--------|
| EJECUCIÓN | N= 20 | N= 30 | N= 40 | N= 100 | N= 20 | N= 30 | N= 40 | N= 100 | N= 20 | N= 30 | N= 40 | N= 100 |
| Mejor | 1,0106 | 1,0106 | 1,0106 | 0,9885 | 1,0637 | 1,0637 | 1,0637 | 1,0637 | 1,0953 | 1,0866 | 1,0106 | 1,0106 |
| Media | 1,4727 | 1,3889 | 1,3392 | 1,1720 | 1,1713 | 1,1328 | 1,1267 | 1,0970 | 1,6809 | 1,5212 | 1,4727 | 1,2915 |
| Peor | 2,4127 | 2,0862 | 1,8881 | 1,6036 | 1,6679 | 1,2242 | 1,2139 | 1,1662 | 2,5549 | 2,4127 | 2,4127 | 1,9262 |
| 1 | 1,0953 | 1,0953 | 1,0953 | 1,0953 | 1,1746 | 1,0972 | 1,1746 | 1,1256 | 1,0953 | 1,0953 | 1,0953 | 1,0953 |
| 2 | 1,2538 | 1,2538 | 1,3252 | 1,1075 | 1,1256 | 1,1662 | 1,0972 | 1,1256 | 1,2242 | 1,2242 | 1,2538 | 1,3031 |
| 3 | 2,2128 | 1,3031 | 1,3031 | 1,1564 | 1,1395 | 1,1256 | 1,1469 | 1,1155 | 1,5699 | 1,2538 | 2,2128 | 1,6236 |
| 4 | 1,3252 | 1,6032 | 1,1075 | 1,1256 | 1,1662 | 1,1075 | 1,1662 | 1,0637 | 1,2538 | 2,2128 | 1,3252 | 1,1075 |
| 5 | 1,3031 | 1,6236 | 1,4189 | 1,3648 | 1,1810 | 1,0853 | 1,1016 | 1,1075 | 2,2128 | 1,3252 | 1,3031 | 1,2680 |
| 6 | 1,6249 | 1,1075 | 1,2680 | 1,1062 | 1,1810 | 1,1075 | 1,0853 | 1,1256 | 2,2660 | 1,3031 | 1,6249 | 1,1564 |
| 7 | 1,6236 | 1,2680 | 1,1564 | 1,0866 | 1,0941 | 1,1256 | 1,1256 | 1,0941 | 1,3252 | 1,6032 | 1,6236 | 1,1256 |
| 8 | 1,1075 | 1,706 4 | 1,2147 | 1,1804 | 1,1469 | 1,1662 | 1,1662 | 1,0853 | 2,3142 | 1,6249 | 1,1075 | 1,2242 |
| 9 | 1,4189 | 1,1564 | 1,1256 | 1,0106 | 1,1395 | 1,1075 | 1,0941 | 1,0637 | 1,3031 | 1,9518 | 1,4189 | 1,3772 |
| 10 | 1,4676 | 1,2147 | 1,2242 | 1,6036 | 1,2242 | 1,2082 | 1,1469 | 1,0853 | 1,6032 | 1,6236 | 1,4676 | 1,3648 |
| 11 | 1,2680 | 1,1256 | 1,3772 | 1,1447 | 1,5621 | 1,2242 | 1,1016 | 1,0637 | 1,6249 | 1,1075 | 1,2680 | 1,1829 |
| 12 | 2,2143 | 1,2555 | 1,6008 | 1,1204 | 1,0941 | 1,2242 | 1,1075 | 1,0637 | 2,2736 | 1,4189 | 2,2143 | 1,1062 |
| 13 | 1,1564 | 1,5515 | 1,1829 | 1,0853 | 1,1469 | 1,0853 | 1,1662 | 1,0637 | 1,9518 | 1,4676 | 1,1564 | 1,0866 |
| 14 | 1,9570 | 1,2242 | 1,2929 | 1,1469 | 1,1075 | 1,1746 | 1,1662 | 1,1075 | 1,6236 | 1,2680 | 1,9570 | 1,0960 |
| 15 | 1,2147 | 1,7745 | 1,1062 | 1,1395 | 1,0941 | 1,0853 | 1,1256 | 1,0853 | 1,6397 | 1,7060 | 1,2147 | 1,2255 |
| 16 | 1,2530 | 1,6008 | 1,0866 | 1,1662 | 1,2923 | 1,0941 | 1,1810 | 1,0853 | 1,1075 | 2,2143 | 1,2530 | 1,1804 |
| 17 | 1,1256 | 1,3648 | 1,0960 | 1,1509 | 1,0853 | 1,1256 | 1,0941 | 1,0853 | 1,4189 | 1,8166 | 1,1256 | 1,0106 |
| 18 | 1,2555 | 1,1829 | 1,5510 | 1,2630 | 1,0853 | 1,1256 | 1,1256 | 1,0853 | 1,4298 | 1,1564 | 1,2555 | 1,3415 |
| 19 | 1,7053 | 1,2929 | 1,2255 | 1,0637 | 1,1469 | 1,0853 | 1,1469 | 1,1155 | 1,4676 | 1,9570 | 1,7053 | 1,9262 |
| 20 | 1,2242 | 1,1062 | 1,1804 | 1,4528 | 1,1972 | 1,1256 | 1,1420 | 1,1155 | 1,9239 | 1,2147 | 1,2242 | 1,6036 |
| 21 | 1,3772 | 1,0866 | 1,0106 | 1,2489 | 1,1395 | 1,0853 | 1,0637 | 1,0853 | 1,2680 | 1,2530 | 1,3772 | 1,1447 |
| 22 | 1,7745 | 2,0604 | 1,3415 | 1,1034 | 1,4442 | 1,1662 | 1,0853 | 1,1155 | 1,7060 | 1,1256 | 1,7745 | 1,5751 |
| 23 | 2,0145 | 1,0960 | 1,5677 | 1,1688 | 1,1075 | 1,1395 | 1,0853 | 1,0637 | 2,2143 | 1,2555 | 2,0145 | 1,1204 |
| 24 | 1,6008 | 1,5510 | 1,6036 | 1,1034 | 1,2198 | 1,0637 | 1,1662 | 1,0853 | 2,5549 | 1,4608 | 1,6008 | 1,4125 |
| 25 | 1,3648 | 1,2255 | 1,8881 | 1,4484 | 1,1075 | 1,1395 | 1,1075 | 1,1256 | 1,8166 | 1,7053 | 1,3648 | 1,6939 |
| 26 | 1,1829 | 1,1810 | 1,1447 | 1,0987 | 1,0637 | 1,0941 | 1,1662 | 1,0941 | 1,1564 | 1,5515 | 1,1829 | 1,0853 |
| 27 | 1,5643 | 1,1804 | 1,5751 | 1,2840 | 1,2242 | 1,2242 | 1,0637 | 1,1662 | 2,0815 | 1,2242 | 1,5643 | 1,1469 |
| 28 | 1,2929 | 1,0106 | 1,3467 | 1,0106 | 1,0941 | 1,0941 | 1,1395 | 1,0941 | 1,9570 | 1,3772 | 1,2929 | 1,2361 |
| 29 | 1,1062 | 1,5711 | 1,1204 | 1,1819 | 1,1256 | 1,1395 | 1,0853 | 1,1016 | 2,1463 | 1,7745 | 1,1062 | 1,1395 |
| 30 | 1,2242 | 1,3415 | 1,4125 | 1,2242 | 1,1810 | 1,0637 | 1,0941 | 1,0853 | 1,2147 | 1,8350 | 1,2242 | 1,1662 |
| 31 | 1,4361 | 1,9401 | 1,7974 | 1,1974 | 1,2242 | 1,1256 | 1,1155 | 1,0941 | 1,2530 | 2,0845 | 1,4361 | 1,3382 |
| 32 | 1,0866 | 1,6036 | 1,0853 | 1,1810 | 1,1420 | 1,1395 | 1,1420 | 1,0941 | 1,6101 | 1,6008 | 1,0866 | 1,1662 |
| 33 | 2,4127 | 1,8881 | 1,3253 | 1,0866 | 1,1256 | 1,0853 | 1,0637 | 1,1395 | 1,1256 | 1,9232 | 2,4127 | 1,1509 |

| | | | | | | | | | | | | |
|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 34 | 1,0960 | 1,5903 | 1,1469 | 1,1856 | 1,1469 | 1,2242 | 1,0853 | 1,1075 | 1,5554 | 1,3648 | 1,0960 | 1,4101 |
| 35 | 1,6726 | 1,1447 | 1,8236 | 1,1395 | 1,1746 | 1,1395 | 1,1662 | 1,0637 | 1,2555 | 1,1829 | 1,6726 | 1,2630 |
| 36 | 1,5510 | 1,5751 | 1,5075 | 1,1256 | 1,0637 | 1,1395 | 1,1395 | 1,0972 | 1,4608 | 1,5643 | 1,5510 | 1,3674 |
| 37 | 1,2255 | 2,0862 | 1,1395 | 1,5624 | 1,1746 | 1,0853 | 1,1155 | 1,0853 | 1,7053 | 1,2929 | 1,2255 | 1,4973 |
| 38 | 1,3521 | 1,1204 | 1,5903 | 1,1892 | 1,1016 | 1,1395 | 1,1155 | 1,1155 | 2,4363 | 1,5967 | 1,3521 | 1,0637 |
| 39 | 1,1810 | 1,4484 | 1,3382 | 0,9885 | 1,0853 | 1,1395 | 1,0637 | 1,0853 | 1,5515 | 1,1062 | 1,1810 | 1,6389 |
| 40 | 1,1804 | 1,4125 | 1,1662 | 1,1395 | 1,1395 | 1,1810 | 1,2082 | 1,0941 | 1,2242 | 1,2242 | 1,1804 | 1,4528 |
| 41 | 1,0106 | 1,8781 | 1,1509 | 1,0187 | 1,0941 | 1,0941 | 1,1395 | 1,1395 | 1,3772 | 1,9213 | 1,0106 | 1,5847 |
| 42 | 1,2716 | 1,3964 | 1,4101 | 1,0187 | 1,1810 | 1,1256 | 1,1469 | 1,0853 | 1,9037 | 1,0866 | 1,2716 | 1,2489 |
| 43 | 1,5711 | 1,0853 | 1,2630 | 1,1804 | 1,1810 | 1,1662 | 1,0941 | 1,0853 | 1,7745 | 2,0604 | 1,5711 | 1,1662 |
| 44 | 1,3415 | 1,3253 | 1,4910 | 1,1544 | 1,1819 | 1,0941 | 1,0972 | 1,0941 | 1,8350 | 2,4127 | 1,3415 | 1,1034 |
| 45 | 1,5677 | 1,1469 | 1,3674 | 1,1810 | 1,1256 | 1,1222 | 1,1256 | 1,0941 | 2,0145 | 1,6921 | 1,5677 | 1,4395 |
| 46 | 1,9401 | 1,2361 | 1,6609 | 1,2476 | 1,2660 | 1,1016 | 1,1256 | 1,1256 | 2,0845 | 1,0960 | 1,9401 | 1,1688 |
| 47 | 1,9262 | 2,0065 | 1,0637 | 1,1586 | 1,2019 | 1,2138 | 1,1395 | 1,0853 | 2,1089 | 1,9531 | 1,9262 | 1,1531 |
| 48 | 1,6036 | 1,5075 | 1,3456 | 1,1358 | 1,0853 | 1,1256 | 1,2139 | 1,0853 | 1,6008 | 1,5510 | 1,6036 | 1,1034 |
| 49 | 1,8932 | 1,1395 | 1,6389 | 1,0195 | 1,1016 | 1,0941 | 1,1819 | 1,1155 | 1,9232 | 1,2255 | 1,8932 | 1,4933 |
| 50 | 1,8881 | 1,1662 | 1,4528 | 1,3571 | 1,6679 | 1,1662 | 1,1256 | 1,0853 | 1,3648 | 1,3521 | 1,8881 | 1,4484 |
| 51 | 1,5903 | 1,3382 | 1,5847 | 1,0637 | 1,1810 | 1,2138 | 1,1395 | 1,0941 | 1,8162 | 1,1810 | 1,5903 | 1,4821 |

6.2.3 Problema S1

| | ALEATORIO | | | | HIPERCUBOS | | | | VECTORES OPUESTOS | | | |
|-----------|-----------|-------|-------|--------|------------|-------|-------|--------|-------------------|-------|-------|--------|
| EJECUCIÓN | N= 20 | N= 30 | N= 40 | N= 100 | N= 20 | N= 30 | N= 40 | N= 100 | N= 20 | N= 30 | N= 40 | N= 100 |
| Mejor | 1,570 | 1,570 | 1,570 | 1,570 | 1,537 | 1,495 | 1,526 | 1,526 | 1,789 | 1,570 | 1,570 | 1,570 |
| Media | 2,474 | 2,341 | 2,222 | 2,091 | 1,606 | 1,578 | 1,572 | 1,557 | 2,841 | 2,637 | 2,482 | 2,482 |
| Peor | 3,872 | 3,663 | 3,065 | 2,545 | 1,727 | 1,664 | 1,605 | 1,582 | 4,119 | 3,872 | 3,872 | 3,872 |
| 1 | 2,147 | 2,144 | 2,144 | 2,144 | 1,579 | 1,605 | 1,590 | 1,543 | 2,147 | 2,147 | 2,147 | 2,147 |
| 2 | 2,144 | 2,839 | 2,342 | 1,972 | 1,573 | 1,574 | 1,582 | 1,543 | 3,242 | 2,144 | 2,144 | 2,144 |
| 3 | 2,839 | 2,342 | 2,005 | 2,049 | 1,572 | 1,571 | 1,583 | 1,540 | 2,144 | 2,839 | 2,839 | 2,839 |
| 4 | 2,342 | 2,005 | 1,972 | 1,789 | 1,582 | 1,584 | 1,591 | 1,559 | 3,215 | 2,881 | 2,342 | 2,342 |
| 5 | 3,530 | 3,125 | 2,068 | 2,419 | 1,645 | 1,543 | 1,590 | 1,551 | 2,839 | 2,342 | 3,530 | 3,530 |
| 6 | 2,005 | 1,972 | 2,069 | 2,514 | 1,586 | 1,602 | 1,574 | 1,545 | 2,881 | 3,181 | 2,005 | 2,005 |
| 7 | 3,349 | 2,245 | 2,049 | 1,570 | 1,590 | 1,581 | 1,580 | 1,548 | 2,342 | 3,461 | 3,349 | 3,349 |
| 8 | 1,972 | 2,069 | 2,608 | 1,869 | 1,727 | 1,576 | 1,587 | 1,543 | 2,748 | 2,005 | 1,972 | 1,972 |
| 9 | 2,068 | 2,049 | 2,382 | 2,091 | 1,573 | 1,607 | 1,543 | 1,566 | 3,816 | 3,349 | 2,068 | 2,068 |
| 10 | 2,245 | 2,608 | 1,789 | 1,858 | 1,574 | 1,642 | 1,575 | 1,574 | 3,530 | 3,125 | 2,245 | 2,245 |
| 11 | 2,069 | 2,382 | 2,419 | 1,993 | 1,718 | 1,574 | 1,553 | 1,574 | 3,449 | 1,972 | 2,069 | 2,069 |
| 12 | 2,348 | 2,412 | 2,421 | 1,936 | 1,597 | 1,573 | 1,573 | 1,547 | 2,005 | 2,068 | 2,348 | 2,348 |
| 13 | 2,049 | 1,789 | 2,431 | 1,801 | 1,569 | 1,620 | 1,574 | 1,543 | 3,349 | 2,245 | 2,049 | 2,049 |
| 14 | 2,541 | 2,419 | 2,661 | 2,179 | 1,708 | 1,543 | 1,554 | 1,574 | 3,889 | 2,485 | 2,541 | 2,541 |
| 15 | 2,608 | 2,678 | 2,558 | 1,960 | 1,584 | 1,571 | 1,579 | 1,565 | 3,125 | 2,069 | 2,608 | 2,608 |
| 16 | 3,872 | 2,421 | 1,570 | 1,986 | 1,578 | 1,664 | 1,526 | 1,543 | 1,972 | 2,348 | 3,872 | 3,872 |
| 17 | 2,382 | 2,431 | 1,822 | 1,879 | 1,599 | 1,575 | 1,553 | 1,571 | 2,068 | 2,049 | 2,382 | 2,382 |
| 18 | 2,882 | 2,513 | 2,532 | 2,102 | 1,569 | 1,571 | 1,543 | 1,545 | 2,535 | 2,137 | 2,882 | 2,882 |
| 19 | 1,789 | 2,558 | 2,007 | 2,294 | 1,582 | 1,549 | 1,570 | 1,549 | 2,245 | 2,608 | 1,789 | 1,789 |
| 20 | 3,020 | 2,989 | 1,869 | 1,953 | 1,588 | 1,570 | 1,572 | 1,582 | 2,485 | 3,217 | 3,020 | 3,020 |
| 21 | 2,419 | 1,570 | 2,091 | 2,249 | 1,711 | 1,581 | 1,545 | 1,543 | 2,905 | 3,872 | 2,419 | 2,419 |
| 22 | 2,678 | 2,496 | 2,131 | 2,050 | 1,596 | 1,575 | 1,573 | 1,573 | 2,069 | 2,382 | 2,678 | 2,678 |
| 23 | 2,421 | 1,822 | 1,858 | 1,727 | 1,621 | 1,579 | 1,574 | 1,579 | 2,878 | 2,412 | 2,421 | 2,421 |
| 24 | 2,444 | 2,532 | 2,294 | 2,344 | 1,590 | 1,574 | 1,551 | 1,558 | 2,348 | 2,882 | 2,444 | 2,444 |
| 25 | 2,431 | 2,007 | 2,253 | 2,223 | 1,580 | 1,589 | 1,588 | 1,570 | 2,049 | 1,789 | 2,431 | 2,431 |
| 26 | 2,513 | 1,869 | 1,993 | 1,863 | 1,576 | 1,587 | 1,570 | 1,552 | 2,137 | 3,020 | 2,513 | 2,513 |
| 27 | 2,661 | 2,227 | 2,405 | 2,179 | 1,607 | 1,495 | 1,573 | 1,528 | 2,541 | 3,537 | 2,661 | 2,661 |
| 28 | 2,881 | 2,091 | 2,188 | 1,947 | 1,625 | 1,571 | 1,588 | 1,552 | 3,395 | 2,419 | 2,881 | 2,881 |

| | | | | | | | | | | | | |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 29 | 2,558 | 2,131 | 1,936 | 2,208 | 1,596 | 1,584 | 1,542 | 1,543 | 2,608 | 2,678 | 2,558 | 2,558 |
| 30 | 3,366 | 2,388 | 2,366 | 2,053 | 1,596 | 1,561 | 1,605 | 1,543 | 3,217 | 2,973 | 3,366 | 3,366 |
| 31 | 1,570 | 1,858 | 2,003 | 2,054 | 1,659 | 1,609 | 1,583 | 1,574 | 4,119 | 2,421 | 1,570 | 1,570 |
| 32 | 2,238 | 3,141 | 1,801 | 1,934 | 1,616 | 1,564 | 1,574 | 1,526 | 3,872 | 2,444 | 2,238 | 2,238 |
| 33 | 2,496 | 2,253 | 2,281 | 2,392 | 1,537 | 1,582 | 1,555 | 1,565 | 2,382 | 2,431 | 2,496 | 2,496 |
| 34 | 1,822 | 1,993 | 2,179 | 2,345 | 1,622 | 1,591 | 1,571 | 1,572 | 2,412 | 3,297 | 1,822 | 1,822 |
| 35 | 3,431 | 2,607 | 2,627 | 1,906 | 1,575 | 1,573 | 1,575 | 1,557 | 3,424 | 2,513 | 3,431 | 3,431 |
| 36 | 2,532 | 2,405 | 3,065 | 2,485 | 1,696 | 1,571 | 1,575 | 1,567 | 2,882 | 2,661 | 2,532 | 2,532 |
| 37 | 2,007 | 2,188 | 1,960 | 2,367 | 1,640 | 1,554 | 1,593 | 1,573 | 1,789 | 2,881 | 2,007 | 2,007 |
| 38 | 3,152 | 1,936 | 1,986 | 2,545 | 1,581 | 1,586 | 1,573 | 1,573 | 3,159 | 2,558 | 3,152 | 3,152 |
| 39 | 1,869 | 2,366 | 2,747 | 2,043 | 1,633 | 1,578 | 1,570 | 1,549 | 3,020 | 2,989 | 1,869 | 1,869 |
| 40 | 2,227 | 2,530 | 2,230 | 1,833 | 1,663 | 1,587 | 1,576 | 1,581 | 3,537 | 3,366 | 2,227 | 2,227 |
| 41 | 2,596 | 2,003 | 2,426 | 2,047 | 1,578 | 1,573 | 1,571 | 1,559 | 2,419 | 1,570 | 2,596 | 2,596 |
| 42 | 2,091 | 1,801 | 1,993 | 1,765 | 1,664 | 1,571 | 1,573 | 1,543 | 3,559 | 2,238 | 2,091 | 2,091 |
| 43 | 2,131 | 2,274 | 1,879 | 2,185 | 1,622 | 1,543 | 1,587 | 1,552 | 2,678 | 2,616 | 2,131 | 2,131 |
| 44 | 2,984 | 3,057 | 2,242 | 2,298 | 1,578 | 1,575 | 1,574 | 1,568 | 3,519 | 2,496 | 2,984 | 2,984 |
| 45 | 2,388 | 2,179 | 2,953 | 2,271 | 1,572 | 1,585 | 1,583 | 1,575 | 2,973 | 1,822 | 2,388 | 2,388 |
| 46 | 1,858 | 3,663 | 2,436 | 2,066 | 1,579 | 1,579 | 1,579 | 1,573 | 2,421 | 3,431 | 1,858 | 1,858 |
| 47 | 2,294 | 2,627 | 2,294 | 1,798 | 1,590 | 1,558 | 1,592 | 1,564 | 2,444 | 2,939 | 2,294 | 2,294 |
| 48 | 3,181 | 3,065 | 2,419 | 2,252 | 1,573 | 1,575 | 1,543 | 1,551 | 2,833 | 2,532 | 3,181 | 3,181 |
| 49 | 2,895 | 1,960 | 1,953 | 2,237 | 1,574 | 1,574 | 1,573 | 1,543 | 3,543 | 2,007 | 2,895 | 2,895 |
| 50 | 2,253 | 2,398 | 2,365 | 2,300 | 1,573 | 1,573 | 1,576 | 1,551 | 2,431 | 3,466 | 2,253 | 2,253 |
| 51 | 1,993 | 1,986 | 2,267 | 2,337 | 1,599 | 1,576 | 1,580 | 1,573 | 3,297 | 3,152 | 1,993 | 1,993 |

6.2.4 Problema YEAST

| | ALEATORIO | | | | HIPERCUBOS | | | | VECTORES OPUESTOS | | | |
|--------------|-----------|-------|-------|--------|------------|-------|-------|--------|-------------------|-------|-------|--------|
| EJECUCIÓN | N= 20 | N= 30 | N= 40 | N= 100 | N= 20 | N= 30 | N= 40 | N= 100 | N= 20 | N= 30 | N= 40 | N= 100 |
| Mejor | 1,626 | 1,626 | 1,117 | 1,117 | 1,041 | 1,041 | 1,041 | 1,041 | 1,758 | 1,626 | 1,626 | 1,117 |
| Media | 2,617 | 2,453 | 2,318 | 2,107 | 1,349 | 1,239 | 1,225 | 1,158 | 2,912 | 2,687 | 2,617 | 2,213 |
| Peor | 3,677 | 3,193 | 3,174 | 3,021 | 1,995 | 1,748 | 1,731 | 1,329 | 4,299 | 3,895 | 3,677 | 2,872 |
| 1 | 2,428 | 2,428 | 2,428 | 1,758 | 1,188 | 1,214 | 1,215 | 1,203 | 3,127 | 3,127 | 2,428 | 1,758 |
| 2 | 2,987 | 1,758 | 1,758 | 1,970 | 1,816 | 1,457 | 1,190 | 1,206 | 2,428 | 2,428 | 2,987 | 2,468 |
| 3 | 1,758 | 2,628 | 2,285 | 2,169 | 1,404 | 1,184 | 1,041 | 1,197 | 3,498 | 1,758 | 1,758 | 1,970 |
| 4 | 3,196 | 2,285 | 1,970 | 2,406 | 1,207 | 1,212 | 1,207 | 1,184 | 2,987 | 2,817 | 3,196 | 2,442 |
| 5 | 2,468 | 1,970 | 2,442 | 2,000 | 1,202 | 1,214 | 1,103 | 1,190 | 1,758 | 3,196 | 2,468 | 2,169 |
| 6 | 2,285 | 2,442 | 2,169 | 2,151 | 1,041 | 1,206 | 1,230 | 1,041 | 2,817 | 2,628 | 2,285 | 2,209 |
| 7 | 2,652 | 2,169 | 2,659 | 2,215 | 1,749 | 1,190 | 1,201 | 1,190 | 3,759 | 2,468 | 2,652 | 2,760 |
| 8 | 1,970 | 2,390 | 2,209 | 1,626 | 1,265 | 1,190 | 1,520 | 1,190 | 3,196 | 2,285 | 1,970 | 2,406 |
| 9 | 2,442 | 2,660 | 2,406 | 2,384 | 1,329 | 1,195 | 1,204 | 1,190 | 2,628 | 2,652 | 2,442 | 2,000 |
| 10 | 3,483 | 2,209 | 2,468 | 2,266 | 1,198 | 1,203 | 1,206 | 1,179 | 2,468 | 1,970 | 3,483 | 2,041 |
| 11 | 2,169 | 2,760 | 2,000 | 1,805 | 1,041 | 1,199 | 1,267 | 1,103 | 2,285 | 2,442 | 2,169 | 2,151 |
| 12 | 2,390 | 2,406 | 2,041 | 2,380 | 1,267 | 1,190 | 1,190 | 1,189 | 3,245 | 2,741 | 2,390 | 2,239 |
| 13 | 2,660 | 2,468 | 2,226 | 2,353 | 1,229 | 1,041 | 1,202 | 1,190 | 2,682 | 3,483 | 2,660 | 2,215 |
| 14 | 2,659 | 3,075 | 2,151 | 2,375 | 1,207 | 1,342 | 1,214 | 1,190 | 2,652 | 2,169 | 2,659 | 2,801 |
| 15 | 2,209 | 2,000 | 2,239 | 2,673 | 1,633 | 1,299 | 1,184 | 1,190 | 1,970 | 2,390 | 2,209 | 1,626 |
| 16 | 2,760 | 2,041 | 2,215 | 2,044 | 1,201 | 1,329 | 1,192 | 1,190 | 2,513 | 3,186 | 2,760 | 2,143 |
| 17 | 3,539 | 2,226 | 2,801 | 2,708 | 1,204 | 1,342 | 1,103 | 1,103 | 2,442 | 2,660 | 3,539 | 2,384 |
| 18 | 2,406 | 2,151 | 1,626 | 1,750 | 1,243 | 1,203 | 1,387 | 1,184 | 2,741 | 3,408 | 2,406 | 2,483 |
| 19 | 2,973 | 3,193 | 2,118 | 1,442 | 1,041 | 1,041 | 1,190 | 1,195 | 3,483 | 2,209 | 2,973 | 2,814 |
| 20 | 2,468 | 2,239 | 2,143 | 1,117 | 1,207 | 1,405 | 1,184 | 1,210 | 4,200 | 2,490 | 2,468 | 2,266 |
| 21 | 3,075 | 2,215 | 2,384 | 2,114 | 1,211 | 1,204 | 1,190 | 1,184 | 2,169 | 2,760 | 3,075 | 1,841 |
| 22 | 2,000 | 2,244 | 2,483 | 1,694 | 1,190 | 1,103 | 1,041 | 1,041 | 4,299 | 2,809 | 2,000 | 1,805 |
| 23 | 2,588 | 2,801 | 2,525 | 1,646 | 1,207 | 1,588 | 1,175 | 1,103 | 2,390 | 2,779 | 2,588 | 2,380 |
| 24 | 2,041 | 1,626 | 3,016 | 1,324 | 1,190 | 1,203 | 1,193 | 1,190 | 3,186 | 2,406 | 2,041 | 2,648 |
| 25 | 2,226 | 2,118 | 2,266 | 1,853 | 1,753 | 1,197 | 1,195 | 1,103 | 2,660 | 2,973 | 2,226 | 2,353 |
| 26 | 3,178 | 2,675 | 1,841 | 2,121 | 1,190 | 1,210 | 1,243 | 1,190 | 3,408 | 2,468 | 3,178 | 2,534 |
| 27 | 2,151 | 2,143 | 1,805 | 2,651 | 1,190 | 1,199 | 1,267 | 1,184 | 3,531 | 3,264 | 2,151 | 2,375 |
| 28 | 3,193 | 2,769 | 2,818 | 2,012 | 1,492 | 1,190 | 1,190 | 1,188 | 2,659 | 3,075 | 3,193 | 2,506 |

| | | | | | | | | | | | | |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 29 | 3,219 | 2,483 | 2,380 | 1,263 | 1,041 | 1,656 | 1,201 | 1,103 | 2,209 | 2,000 | 3,219 | 2,872 |
| 30 | 2,239 | 2,525 | 2,648 | 2,296 | 1,203 | 1,202 | 1,329 | 1,190 | 3,661 | 2,588 | 2,239 | 2,673 |
| 31 | 2,215 | 2,814 | 2,353 | 2,403 | 1,281 | 1,190 | 1,387 | 1,190 | 2,760 | 2,695 | 2,215 | 2,044 |
| 32 | 2,244 | 3,016 | 2,701 | 2,236 | 1,041 | 1,208 | 1,707 | 1,190 | 2,809 | 2,041 | 2,244 | 2,263 |
| 33 | 3,677 | 2,266 | 2,534 | 1,847 | 1,733 | 1,041 | 1,201 | 1,190 | 3,539 | 2,898 | 3,677 | 2,708 |
| 34 | 2,801 | 2,756 | 2,375 | 2,516 | 1,611 | 1,204 | 1,204 | 1,041 | 3,583 | 2,226 | 2,801 | 2,854 |
| 35 | 3,466 | 1,841 | 2,635 | 2,165 | 1,190 | 1,103 | 1,103 | 1,041 | 2,779 | 3,178 | 3,466 | 1,750 |
| 36 | 1,626 | 1,805 | 2,872 | 2,412 | 1,206 | 1,287 | 1,190 | 1,041 | 2,406 | 2,151 | 1,626 | 2,807 |
| 37 | 2,118 | 2,820 | 2,673 | 1,939 | 1,231 | 1,538 | 1,210 | 1,190 | 3,222 | 3,193 | 2,118 | 1,442 |
| 38 | 2,875 | 2,380 | 2,044 | 1,685 | 1,299 | 1,103 | 1,184 | 1,329 | 2,973 | 3,524 | 2,875 | 2,387 |
| 39 | 2,675 | 2,973 | 2,932 | 2,163 | 1,678 | 1,188 | 1,211 | 1,191 | 2,468 | 3,219 | 2,675 | 1,117 |
| 40 | 2,143 | 2,648 | 2,263 | 1,780 | 1,902 | 1,184 | 1,190 | 1,207 | 3,264 | 2,239 | 2,143 | 2,007 |
| 41 | 2,384 | 2,879 | 2,708 | 2,021 | 1,616 | 1,190 | 1,208 | 1,203 | 3,581 | 2,215 | 2,384 | 2,290 |
| 42 | 2,836 | 2,353 | 2,854 | 2,128 | 1,528 | 1,329 | 1,190 | 1,184 | 3,075 | 2,592 | 2,836 | 2,114 |
| 43 | 2,500 | 2,701 | 2,020 | 2,305 | 1,212 | 1,230 | 1,206 | 1,175 | 3,703 | 2,244 | 2,500 | 1,694 |
| 44 | 2,483 | 2,534 | 1,750 | 2,453 | 1,995 | 1,190 | 1,103 | 1,103 | 2,000 | 3,677 | 2,483 | 2,275 |
| 45 | 2,525 | 2,375 | 2,807 | 2,525 | 1,907 | 1,748 | 1,214 | 1,041 | 2,588 | 3,051 | 2,525 | 1,871 |
| 46 | 2,814 | 2,506 | 1,442 | 2,470 | 1,610 | 1,103 | 1,190 | 1,190 | 3,399 | 2,801 | 2,814 | 1,646 |
| 47 | 3,477 | 2,969 | 3,174 | 2,636 | 1,041 | 1,203 | 1,202 | 1,190 | 2,695 | 3,895 | 3,477 | 1,324 |
| 48 | 3,016 | 2,872 | 1,117 | 1,608 | 1,635 | 1,230 | 1,731 | 1,197 | 2,041 | 1,626 | 3,016 | 2,523 |
| 49 | 2,781 | 2,673 | 2,155 | 2,211 | 1,195 | 1,208 | 1,192 | 1,041 | 2,898 | 2,118 | 2,781 | 2,417 |
| 50 | 2,266 | 2,780 | 2,007 | 3,021 | 1,369 | 1,103 | 1,203 | 1,103 | 2,226 | 2,875 | 2,266 | 1,853 |
| 51 | 2,756 | 2,044 | 2,290 | 2,365 | 1,190 | 1,214 | 1,190 | 1,041 | 3,439 | 2,945 | 2,756 | 2,121 |

6.3 Cuestionario a farmacéuticos de Sevilla

LA FARMACIA

Elegir una única opción

1. Su tipo de farmacia es
 - ☐ Rural
 - ☐ Urbana
2. ¿Su farmacia es la única en la localidad?
 - ☐ Si
 - ☐ No
3. ¿La distancia al centro de salud más próximo es menor a 250 metros?
 - ☐ Si
 - ☐ No

COMUNICACIÓN CON CLIENTES/PACIENTES

4. ¿Distingue entre clientes y pacientes?
 - ☐ Si
 - ☐ No
5. ¿Qué medios utiliza para comunicarse con sus clientes/pacientes

| | Si | No | 6. ¿Cuál es el que más utiliza? |
|--|----|----|---------------------------------|
| Correo electrónico | | | |
| SMS (mensajes de móvil) | | | |
| Teléfono | | | |
| Aplicación de mensajería móvil (ej whatsapp) | | | |
| Redes sociales | | | |
| Blog | | | |
| Twitter | | | |
| Facebook | | | |
| Otros | | | |

CONOCIMIENTO DE CLIENTES/PACIENTES

7. ¿Posee su farmacia Tarjeta de Fidelización?
 - ☐ Si
 - ☐ No

| 8. ¿Qué información registra de sus clientes/pacientes? | Si | No |
|---|----|----|
| Frecuencia de compra/visita | | |
| Hábitos de consumo (tabaco, alcohol,...) | | |
| Preferencia en servicios (seguimiento de la medicación, dietas,...) | | |
| Hábitos para el cuidado de la salud (deporte, control medicación,...) | | |
| Temas de interés para los pacientes/clientes (dermocosmética, cuidado pediátrico, ...) | | |
| Otra información relacionada con el conocimiento de los pacientes/clientes (indicar cuál) | | |

6.4 Código C++ del algoritmo DECCS utilizado en la investigación

//DECCS (Differential Evolution for Clustering with Continuation Scheme)

// main.cpp

/*Librerías*/

#include <iostream>

#include <fstream>

#include <cmath>

#include <vector>

#include <sstream>

#include <string>

#include <time.h>

using namespace std;

/*Dimensionado del problema*/

#define dimvar 350

#define dimpob 610 //20 para DECCS y 10*variables para ACDE

#define dimcolumns 60

#define dimCluster 30

#define dimrows 14000

#define dimrun 51

/*Clases*/

class Cluster{

public:

bool active;

int nk; //Tamaño del clúster

double center[dimcolumns]; //Centro del cluster indicado por x[nvar]

double centroide[dimcolumns]; //centroide (d1,d2,...,dm)

};

```

class Punto{
public:
    void Evaluar();
    void GenerarPunto();
    void CalcularActiveClusters();
    void AsignarPuntosACluster();
    void GuardarCenter(int cluster);
    void CalcularCentroides();
    double CS(); //Indicador de CS
    double DN(); //Indicador de Dunn
    double CH(); //Indicador de Calinski-Harabasz
    double DB(); //Indicador de Davies-Bouldin
    void Ordenar_X();
public:
    double x[dimvar];
    double f;
    int numclusters; //Número de clústeres activos
    int asign[dimrows]; //cluster al que pertenece
    double between_dist[dimCluster][dimCluster]; //distancia entre clústeres
    double between_min[dimCluster]; //distancia mínima entre clústeres
    double within_max[dimrows];
    double within_mean[dimCluster];
    double num,den;
};

class Arranque{ /*Clase necesaria para los hipercubos latinos*/
public:
    int Uno[dimvar];
};

/*Variables globales*/

```

```
int Kmax;    //maximum number of addmitted clusters
int nvar;    //Variables para la codificación
int columns; //Características-variables del problema
int rows;    //Filas de la matriz de datos (vectores)
int tampob;  //Tamaño de la población
int iter;    //Número de iteraciones
double cotasup[dimcolumns], cotainf[dimcolumns]; //Cotas de las variables
double FalseMax[dimvar]; //Esquema de continuación
double graph[dimvar][5000]; //para guardar el CDF(var) del esquema de cont.
int tamx[dimvar]; //variable auxiliar para grap
double pc; //probabilidad de cruce de la ED
double F; //Inercia de la ED
double HMR; //High Mutation Ratio del esquema de cont.
int contScheme; //Var. Binaria para usar o no el esquema de cont.
int combined; //Variable entera para usar cruce bin, exp o ambos
int eval; //Número de evaluaiones de la FO usadas
int run; //Número de ejecuciones a realizar
int semillado; //variable entera para el tipo de semillado usado
vector <double> data; //datos (en vector) del problema
double matrix[dimrows][dimcolumns]; //copio data en matrix (en una matriz)
double centro_datos[dimrows]; //valores medios por columnas
double distmatrix[dimrows][dimrows]; //matriz de distancias
Punto Pob[dimpob]; //Población de la ED
Cluster CL[dimCluster]; //Clústeres máximo a usar
Cluster CL_opt[dimCluster]; //Clúster óptimo
Arranque Ini[1000]; //Vector auxiliar para los hipercubos latinos
/*Los siguientes vectores alamacenan información de todas las evaluaciones para
posteriormente analizarlas estadísticamente*/
double run_f[dimrun];
double run_k[dimrun];
double run_intra[dimrun];
double run_entre[dimrun];
```

```

double run_time[dimrun];
double run_eval[dimrun];
double run_velocity[4000][dimrun];
clock_t tini, tfin;    //Tiempo de inicio y finalización de una ejecución

char name_pob[150] = "DatosA.txt";    //PROBLEMA A RESOLVER

int ReadNumbers( const string & s, vector <double> & v );
void import_matrix_from_txt_file(const char* filename_X, vector <double>& v, int&
rows, int& cols);
double SquareEuclideanDist(double p1[],double p2[], int lenght);

int main(int argc, const char * argv[]) //Función principal
{
    void LeerParametrosDE();
    void generarDE(int iterac);
    void SalidaDatos(int );
    void GuardarMultipleRun();
    void LeerProblema();
    void generarPopInicial(int sem);

    import_matrix_from_txt_file(name_pob,data,rows,columns);
    LeerParametrosDE();

    for (int r=0; r<run; r++) {    /*Resolvemos el problema run vece*/
        tini=clock();
        eval=0;

        cout<<"Run: "<<r<<endl;
        generarPopInicial(semillado);

        for (int i=0; i<iter; i++) {    /*Iteraciones de la ED*/

```

```
    generarDE(i);
    run_velocity[i][r] = Pob[0].f;

    tfin = clock();
    if (difftime(tfin, tini)/CLOCKS_PER_SEC > 3600)
        i+=iter;
}

tfin = clock();

//Datos para las múltiples ejecuciones
run_time[r] = difftime(tfin, tini)/CLOCKS_PER_SEC;
run_f[r] = Pob[0].f;
run_k[r] = Pob[0].numclusters;
run_intra[r] = Pob[0].num;
run_entre[r] = Pob[0].den;
run_eval[r] = eval;

SalidaDatos(r);          //Guardamos información de la ejecución "r"
}

GuardarMultipleRun();    //Guardamos en un fichero información
                          estadística de las r ejecuciones

return 0;
}

void LeerParametrosDE()
{
    tampob    = 60;      //20 para DECCS y 10*NumVar para ACDE
    run       = 40;      //Ejecuciones
    iter      = 650;     //Iteraciones a realizar por la ED
    pc        = 0.5;
```

```

F      = 1;

contScheme = 1;  //0= Sin esquema de cont; 1 CON esquema de cont.
combined   = 2;  //0= solo bin, 1 = solo exp, 2 = combined
HMR        = 0.04;  //High Mutation Ratio para el esquema de cont.

semillado  = 1;  //0=naive;1=Latyn;2=Opossited-Based
Kmax = 10;
nvar = Kmax + Kmax*columns;

for (int i=0; i<nvar; i++) {
    if (i < Kmax)
        FalseMax[i] = 1/5.;
    else
        FalseMax[i] = (cotasup[(i - Kmax)%columns] - cotainf[(i - Kmax)%columns])/5.;
}

if (tampob >= dimpob )
    cout<<"Error con el dimensionado de dimpob"<<endl;

if (Kmax > dimCluster)
    cout<<"Ojo con Kmax y dimcluster!!!"<<endl;

if (nvar > dimvar)
    cout<<"Dimensionar problema. Nvar excede dimvar!!!"<<endl;
}

void generarPopInicial(int seed)
{
    void Naive();
    void Latyn();
    void OBased();
}

```

```
if(seed == 0)
    Naive();
else if(seed == 1)
    Latyn();
else if(seed == 2)
    OBased();
}

void Naive()
{
    Pob[0].f = 1E10;
    eval = 0;

    for (int i=0; i<tampob; i++) {
        Pob[i].GenerarPunto();
        Pob[i].Evaluar();

        if (Pob[i].f < Pob[0].f) {
            Pob[0]=Pob[i];

            for (int j=0; j<Kmax; j++)
                CL_opt[j]=CL[j];

            cout<<"Mejora: "<<Pob[0].f<<endl;
        }
    }
}

void Latyn()
{
    void InicializarArranques(int tam);
    void GenerarPuntos(int tampob);
```

```
InicializarArranques(tampob);

int i;
int var,sol1,sol2,aux;

for(i=0;i<1000000;i++) //Permutaciones aleatorias
{
    var = rand()%nvar;
    sol1 = rand()%tampob;
    do{
        sol2 = rand()%tampob;
    }
    while(sol2==sol1);

    aux = Ini[sol1].Uno[var];
    Ini[sol1].Uno[var] = Ini[sol2].Uno[var];
    Ini[sol2].Uno[var] = aux;
}

GenerarPuntos(tampob);
}

void InicializarArranques(int tam)
{
    int i,j;

    for(i=0;i<tam;i++)
        for(j=0;j<nvar;j++)
            Ini[i].Uno[j] = i;
}

void GenerarPuntos(int tam)
{
```



```
int i,j;
double prob;
double salto;

for(i=0;i<tam;i++)
{
    for(j=0;j<nvar;j++)
    {
        if (j < Kmax) { //Las primeras variables entre 0 y 1

            prob = 0.5; // Puntos medios de los cubos
            salto = 1.0/tam;

            Pob[i].x[j] = 0 + salto*(Ini[i].Uno[j] + prob);
        }
        else{ //El resto de variables entre cotainf y cotasup

            prob = 0.5;
            int col = (j-Kmax)%columns;
            salto = (cotasup[col]-cotainf[col])/tam;

            Pob[i].x[j] = cotainf[col] + salto*(Ini[i].Uno[j] + prob);
        }
    }
}

int active=0;
for (j=0; j<Kmax; j++)
    if (Pob[i].x[j]>0.5)
        active++;
if (active < 2){
    Pob[i].x[0]=0.75;
    Pob[i].x[1]=0.75;
```

```
    }

    Pob[i].Evaluar();

    if (Pob[i].f < Pob[0].f) {
        Pob[0]=Pob[i];

        for (int j=0; j<Kmax; j++)
            CL_opt[j]=CL[j];

        cout<<"Mejora: "<<Pob[0].f<<endl;
    }
}

void OBased()
{
    Pob[0].f = 1E10;
    eval = 0;

    for (int i=0; i<tampob/2; i++) {
        Pob[i].GenerarPunto();
        Pob[i].Evaluar();

        if (Pob[i].f < Pob[0].f) {
            Pob[0]=Pob[i];

            for (int j=0; j<Kmax; j++)
                CL_opt[j]=CL[j];

            cout<<"Mejora: "<<Pob[0].f<<endl;
        }
    }
}
```

```
for (int i=0; i<tampob/2; i++) {
    for (int j=0; j<nvar; j++)
    {
        if (j<Kmax) {
            Pob[tampob/2+i].x[j] = 1-Pob[i].x[j];
        }
        else
            Pob[tampob/2+i].x[(j-Kmax)%columns] = cotainf[(j-
Kmax)%columns]+cotasup[(j-Kmax)%columns]-Pob[i].x[(j-Kmax)%columns];
    }

    Pob[tampob/2+i].Evaluar();

    if (Pob[tampob/2+i].f < Pob[0].f) {
        Pob[0]=Pob[tampob/2+i];

        for (int j=0; j<Kmax; j++)
            CL_opt[j]=CL[j];

        cout<<"Mejora: "<<Pob[0].f<<endl;
    }
}
}
```

```
double F_inertia(double i)
{
    double val;
    if (i==0.5)
        val = 0.5;
    else
        val = 0.5 +rand() /((double)RAND_MAX+1);
}
```

```

    return val;
}

double CR_variable(int ii)
{
    double cr;

    cr = pc +(1.-pc)*(iter-ii)/((double) iter);

    return cr;
}

void generarDE(int iteracion)
{
    int flip(double p);
    int r1,r2,r3;
    double aux;
    int elegidavar; //solo aplicamos esquema de continuación CDF() en 1 variable
    int exito;
    int usado_per;

    double F_inertia(double );
    double CR_variable(int );
    int ExponentialDist(int );
    double CDF(int var);
    double Pertubar(int var);

    int i,k;
    for (i = 0; i < tampob; i++)
    {
        do // Se selecciona un primer padre
            r1 = (int) ((rand() /((double)RAND_MAX+1))* tampob);
        while(r1 == i);
    }

```

```
do // Se selecciona un segundo padre
    r2 = (int) ((rand() / ((double)RAND_MAX+1)) * tampob);
while(r2 == i || r2 == r1);
do // Se selecciona un tercer padre
    r3 = (int) ((rand() / ((double)RAND_MAX+1)) * tampob);
while(r3 == i || r3 == r2 || r3 == r1 );

int var_aux = (int) ((rand() / ((double)RAND_MAX+1)) * nvar);

elegidavar = 0; //solo aplicamos esquema de cont. en 1 variable
exito=0;
usado_per=-1;

if((combined==0) || ( (combined == 2)&&(rand() / ((double)RAND_MAX+1)) < 0.5) )
//BIN crossover
{
    for (k = 0; k < nvar; k++)
    {
        if ((rand() / ((double)RAND_MAX+1)) < CR_variable(iteracion) || k == var_aux)
        {
            if ((contScheme == 1)&&(elegidavar==0)&&(k >= Kmax)) {
                aux = Pob[r3].x[k] + F * CDF(k);
                elegidavar = 1;
            }
            else
                aux = Pob[r3].x[k] + F_inertia(F) * (Pob[r2].x[k] - Pob[r1].x[k]);
        }
        else
            aux = Pob[i].x[k];
    }

    // Aqui comprobamos que no se salga del límite
    if(k < Kmax){
```

```

    if (aux < 0)
        aux = 0.5*(rand() /((double)RAND_MAX+1));
    else if(aux>1)
        aux=0.5 + 0.5*(rand() /((double)RAND_MAX+1));

    Pob[tampob].x[k] = aux;
}
else {
    if ( (aux < cotainf[(k-Kmax)%columns]) || (aux>cotasup[(k-Kmax)%columns]))
        aux = cotainf[(k-Kmax)%columns] + (rand()
/((double)RAND_MAX+1))*(cotasup[(k-Kmax)%columns]-cotainf[(k-Kmax)%columns]);

    Pob[tampob].x[k] = aux;
}
}
}
else if((combined==1) || (combined==2)) //EXP crossover
{
    int var_aux2 =(int) ((rand() /((double)RAND_MAX+1)) * nvar);
    int L = ExponentialDist(var_aux2+1);

    for (k = 0; k < nvar; k++)
    {
        if ( (var_aux2 + L < nvar) && (var_aux2 <= k) && (k <= (L+var_aux2)%nvar))
        {
            if( (L==1) && (k >= Kmax) && (rand() /((double)RAND_MAX+1)) < HMR)
//high mutation ratio scheme
            {
                aux = Pob[r3].x[k] + F * Pertubar(k);
                usado_per = k;
            }
        }
        else{

```

```
if ((contScheme == 1)&&(elegidavar==0)&&(k >= Kmax)) {
    aux = Pob[r3].x[k] + F * CDF(k);
    elegidavar=1;
}
else
    aux = Pob[r3].x[k] + F_inertia(F) * (Pob[r2].x[k] - Pob[r1].x[k]);
}
}
else if ( (var_aux2 + L >= nvar) && ((var_aux2 >= k) || (k <= (L +
var_aux2)%nvar)) )
{
    if((L==1) && (k >= Kmax) && (rand() /((double)RAND_MAX+1)) < HMR) //high
mutation scheme
    {
        aux = Pob[r3].x[k] + F * Pertubar(k);
        usado_per=k;
    }else
    {
        if ((contScheme == 1)&&(elegidavar==0)&&(k >= Kmax)) {
            aux = Pob[r3].x[k] + F * CDF(k);
            elegidavar = k;
        }
        else
            aux = Pob[r3].x[k] + F_inertia(F) * (Pob[r2].x[k] - Pob[r1].x[k]);
        }
    }
}
else
    aux = Pob[i].x[k];

// Aqui comprobamos que no se salga del límite
if(k < Kmax){
    if (aux < 0)
```

```

        aux = 0.5*(rand() /((double)RAND_MAX+1));
    else if(aux>1)
        aux=0.5 + 0.5*(rand() /((double)RAND_MAX+1));

    Pob[tampob].x[k] = aux;
}
else {
    if ( (aux < cotainf[(k-Kmax)%columns]) || (aux>cotasup[(k-Kmax)%columns]))
        aux = cotainf[(k-Kmax)%columns] + (rand() /((double)RAND_MAX+1))*(cotasup[(k-Kmax)%columns]-cotainf[(k-Kmax)%columns]);

    Pob[tampob].x[k] = aux;
}
}
}

// Evaluo la aptitud del individuo generado por los 3 padres
int active=0;
for (k=0; k<Kmax; k++)
    if (Pob[tampob].x[k]>0.5)
        active++;

if (active < 2){ //Forzamos que haya al menos dos clúster activos
    Pob[tampob].x[0]=0.75;
    Pob[tampob].x[1]=0.75;
}

Pob[tampob].Evaluar(); //Evaluamos la solucion
if (Pob[tampob].f < Pob[i].f) {
    Pob[i]=Pob[tampob];

    if(usado_per>=0)

```



```
FalseMax[usado_per] += 0.1*(FalseMax[usado_per] -
graph[usado_per][tamx[usado_per] - 1]);

if (Pob[tampob].f < Pob[0].f) {

    Pob[0]=Pob[tampob];

    for (int j=0; j<Kmax; j++)
        CL_opt[j]=CL[j];

    cout<<"Mejora: "<<Pob[0].f<<" con clusters: "<<Pob[0].numclusters<<endl;
}
}
else{
    if(usado_per>=0)
        FalseMax[usado_per] -= 0.1*(FalseMax[usado_per]-
graph[usado_per][tamx[usado_per]-1]);
    }
}
}

double CDF(int selectedvar) /*Calculamos la CDF del esquema de cont*/
{
    double xvalue = 0.0;
    int i,j;

    tamx[selectedvar]=0;
    for (i=0; i<tampob-1; i++) {
        for (j=i+1; j<tampob; j++)
        {
            graph[selectedvar][tamx[selectedvar]] = abs(Pob[i].x[selectedvar]-
Pob[j].x[selectedvar]);
            tamx[selectedvar]++;
        }
    }
}
```

```

    }
}

```

```

//Ordenamos la lista graph[]

```

```

for (i=0; i<tamx[selectedvar]; i++) {
    for (j=i+1; j<tamx[selectedvar]; j++) {
        if (graph[selectedvar][i] > graph[selectedvar][j]) {
            double aux;
            aux = graph[selectedvar][i];
            graph[selectedvar][i] = graph[selectedvar][j];
            graph[selectedvar][j] = aux;
        }
    }
}

```

```

//Seleccionamos un valor aleatorio entre 0 y 1 y CDF-1

```

```

double pvalue = rand()/((double) RAND_MAX + 1);
int suelo = floor(pvalue*(tampob*(tampob-1)/2-1));

```

```

xvalue = graph[selectedvar][suelo] + (graph[selectedvar][suelo + 1] -
graph[selectedvar][suelo])*( pvalue*(tampob*(tampob-1)/2-1) - suelo);

```

```

if ( rand()/((double)RAND_MAX+1) < 0.5)//Devolvemos el valor obtenido

```

```

    return xvalue;

```

```

else

```

```

    return -xvalue;

```

```

}

```

```

double Pertubar(int selectedvar) /*Generamos valor con FalseMax*/

```

```

{

```

```

    double xvalue = 0.0;

```

```
int i,j;

tamx[selectedvar]=0;
for (i=0; i<tampob-1; i++) {
    for (j=i+1; j<tampob; j++) {
        if (Pob[i].x[selectedvar]-Pob[j].x[selectedvar] > 0) {
            graph[selectedvar][tamx[selectedvar]] = Pob[i].x[selectedvar]-
Pob[j].x[selectedvar];
            tamx[selectedvar]++;
        }
    }
}
```

//Ordenamos la lista graph[]

```
for (i=0; i<tamx[selectedvar]; i++) {
    for (j=i+1; j<tamx[selectedvar]; j++) {
        if (graph[selectedvar][i] > graph[selectedvar][j]) {
            double aux;
            aux = graph[selectedvar][i];
            graph[selectedvar][i] = graph[selectedvar][j];
            graph[selectedvar][j] = aux;
        }
    }
}
```

//Seleccionamos un valor aleatorio entre graph[tamx-1] y FalseMax[var_i]

```
if(FalseMax[selectedvar]-graph[selectedvar][tamx[selectedvar]-1]>0)
    xvalue = graph[selectedvar][tamx[selectedvar]-1] + (FalseMax[selectedvar]-
graph[selectedvar][tamx[selectedvar]-1]) * (rand()/((double) RAND_MAX +1));
else
    xvalue = graph[selectedvar][tamx[selectedvar]-1];
```

```
    if (rand()/((double) RAND_MAX + 1) < 0.5)
        return xvalue;
    else
        return -xvalue;

}

int ExponentialDist(int nmax)
{
    int entero = 1;

    double aleatorio = rand() / ((double) RAND_MAX + 1);
    double distmin = 1000000;

    for (int i=1; i < nmax; i++)
        if (abs((1-pc)*pow(pc, i-1) - aleatorio) < distmin) {
            entero = i;
            distmin = abs((1-pc)*pow(pc, i-1) - aleatorio);
        }

    if (abs(pow(pc, nmax-1) - aleatorio) < distmin)
        entero = nmax;

    return entero;
}

void Punto::GenerarPunto()
{
    int i,j;
    for (i=0; i<Kmax; i++)
        x[i] = (rand()%10001)/(10000.0);
```

```
int active=0;
for (i=0; i<Kmax; i++)
    if (x[i]>0.5)
        active++;
if (active < 2){
    x[0]=0.75;
    x[1]=0.75;
}
for (i=0; i<Kmax; i++)
    for (j=0; j<columns; j++) {
        x[Kmax + i * columns + j] = cotainf[j] + (rand()%10001)/(10000.0) * (cotasup[j] -
cotainf[j]);
    }
}

void Punto::Evaluar() /*Función objetivo*/
{
    eval++;

    //Asignar centers de x[dimvar] a CL[i].center[dimvar]
    int i;
    for (i=0; i<Kmax; i++)
        GuardarCenter(i);

    //Calcular número de clúster activos ** CL[i].active **
    CalcularActiveClusters();

    //Asignar puntos a clústeres según distancia al center point del array de variables **
    bool asign[dimrows], almacenar tamaño de cada cluster ** int CL[i].nk[Kmax]
    AsignarPuntosACluster();

    //Calcular centroides para cada clúster ** double centroide[columns]
```

```

CalcularCentroides();

//Distancias entre centroides, almacenar el min ** between_dist[columns][columns]
double between_min
if (numclusters==1) {
    f=100000000;
}
else{
    f = CS();    //Indicador de CS
    //f = DN(); //Indicador de Dunn
    //f = CH(); //Indicador de Calinski-Harabasz
    //f = DB(); //Indicador de Davies-Bouldin
}

//Esquema de reordenamiento para evitar simetría en las soluciones
Ordenar_X();
}

// Función que hace un flip con el porcentaje recibido
int flip (double pm)
{
    double ran = (rand() /((double)RAND_MAX+1));
    if ( ran <= pm)
        return 1;
    else
        return 0;
}

void GuardarMultipleRun()    /*Estadísticas de las ejecuciones*/
{
    double mean(double v[], int d);
    double desvestp(double v[],int d);

```

```
double max(double v[], int d);
double min(double v[], int d);

char Sal2[150];
int i, j;

sprintf(Sal2,"ACDE_%s_allrun_CH.txt",name_pob);
//sprintf(Sal2,"DECCS_%s_allrun.txt",name_pob);

filebuf bufExc2;
if ( bufExc2.open(Sal2 , ios::out ) == NULL )
{
    cerr<<"Error: no se puede abrir el fichero 1\n";
    exit(1);
}
ostream salExc2( &bufExc2);

salExc2 <<"Run " <<"fobj " <<"K " <<"Intra " <<"Entre " <<"Eval " <<"Time"<<endl;
for (i=0; i<run; i++) {
    salExc2  <<i<<"      " <<run_f[i]<<"      " <<run_k[i]<<"      " <<run_intra[i]<<"
    "<<run_entre[i]<<" "<<run_eval[i]<<" "<<run_time[i]<<endl;
}

salExc2 <<endl;

salExc2  <<"Min"<<"      " <<min(run_f,run)<<"      " <<min(run_k,run)<<"
"<<min(run_intra,run)<<"      " <<min(run_entre,run)<<"      " <<min(run_eval,run)<<"
"<<min(run_time,run)<<endl;

salExc2  <<"Mean"<<"      " <<mean(run_f,run)<<"      " <<mean(run_k,run)<<"
"<<mean(run_intra,run)<<"      " <<mean(run_entre,run)<<"      " <<mean(run_eval,run)<<"
"<<mean(run_time,run)<<endl;

salExc2  <<"Max"<<"      " <<max(run_f,run)<<"      " <<max(run_k,run)<<"
"<<max(run_intra,run)<<"      " <<max(run_entre,run)<<"      " <<max(run_eval,run)<<"
"<<max(run_time,run)<<endl;
```

```

    salExc2 <<"StdDev"<<"    "<<desvestp(run_f,run)<<"    "<<desvestp(run_k,run)<<"
"<<desvestp(run_intra,run)<<"    "<<desvestp(run_entre,run)<<"
"<<desvestp(run_eval,run)<<" "<<desvestp(run_time,run)<<endl;

```

```

salExc2 <<endl;

```

```

salExc2 <<"Iter"<<" "<<"Min"<<" "<<"Mean"<<" "<<"Max"<<" "<<"SD"<<endl;

```

```

//Velocidad de convergencia de todas las ejecuciones

```

```

for (int it=0; it<iter; it++) {
    salExc2<<it<<" "<<min(run_velocity[it],run)<<" "<<mean(run_velocity[it],run)<<"
"<<max(run_velocity[it],run)<<" "<<desvestp(run_velocity[it],run)<<endl;
}
}

```

```

void SalidaDatos(int rr)    /*Óptimo de la ejecución rr-ésima*/

```

```

{
    char Sal[150];
    int i, j;

    sprintf(Sal,"ACDE_%s_run_%d_CH.txt",name_pob,rr);
    //sprintf(Sal,"DECCS_%s_run_%d.txt",name_pob,rr);

    filebuf bufExc;
    if ( bufExc.open(Sal , ios::out ) == NULL )
    {
        cerr<<"Error: no se puede abrir el fichero 1\n";
        exit(1);
    }
    ostream salExc( &bufExc);

    cout<<"Óptimo con f= :"<<Pob[0].f<<" tras evaluaciones: "<<eval<<endl;
    cout<<" Mandando a fichero..."<<endl;

```



```
salExc <<"Optimo: " <<Pob[0].f<<endl;
salExc<<"Evaluaciones: " <<eval<<endl;
salExc<<"Clusters: " <<Pob[0].numclusters<<endl;

salExc<<"Tamaño y centroides clusters encontrados:"<<endl;
for (i=0; i<Kmax; i++){

    salExc<<CL_opt[i].nk<<" ";

    for (j=0; j<columns; j++) {
        salExc<<CL_opt[i].centroide[j]<<" ";
    }
    salExc<<endl;
}

salExc<<endl;

salExc<<"Distancia entre cluster"<<endl;
for (i=0; i<Kmax; i++)
{
    for (j=0; j<Kmax; j++){
        salExc<<Pob[0].between_dist[i][j]<<" "; //distancia entre clústeres
    }
    salExc<<endl;
}
salExc<<endl;

salExc<<"Distancia intra-cluster"<<endl;

for (i=0; i<Kmax; i++)
    salExc<<Pob[0].within_mean[i]<<" "; //distancia entre clústeres
```

```
salExc<<endl;

for (i=0; i<rows; i++) {
    salExc<<Pob[0].assign[i]<<endl;
}

}

int ReadNumbers( const string & s, vector <double> & v ) {
    istringstream is( s );
    double n;
    while( is >> n ) {
        v.push_back( n );
    }
    return v.size();
}

void import_matrix_from_txt_file(const char* filename_X, vector <double>& v, int&
row, int& cols){

    ifstream file_X;
    string line;

    file_X.open(filename_X);
    if (file_X.is_open())
    {
        int i=0;
        getline(file_X, line);
        cols =ReadNumbers( line, v );
        cout << "cols:" << cols << endl;
        for ( i=1;i<32767;i++)
```

```
{
    if ( !getline(file_X, line) ) break;
    ReadNumbers( line, v );
}

row=i;
cout << "rows :" << row << endl;
if(row >32766) cout<< "N must be smaller than MAX_INT";

file_X.close();
}
else{
    cout << "file open failed";
}
```

//Copiamos v en matrix

```
for (int i=0;i<row;i++)
    for (int j=0;j<cols;j++)
        matrix[i][j]=v[ i*cols + j ];
```

//Calculamos las cotas inf y sup de las variables

```
for (int j=0;j<cols;j++)
{
    cotainf[j] = 10000000000;
    cotasup[j] = -10000000000;

    for (int i=0;i<row;i++)
    {
        if(cotainf[j] > v[i*cols+j])
            cotainf[j] = v[i*cols+j];
        else if(cotasup[j] < v[i*cols+j])
            cotasup[j] = v[i*cols+j];
    }
}
```

```

    }
}

//Calculamos la matriz de distancias entre todos los puntos de matrix
for (int i=0;i<row;i++)
    for (int j=i+1;j<row;j++)
    {
        distmatrix[i][j] = SquareEuclideanDist(matrix[i], matrix[j], cols);
        distmatrix[j][i] = distmatrix[i][j];
    }

for (int j=0;j<cols;j++)
    cout<<"column  "<<j<<" from "<<cotainf[j]<<" to "<<cotasup[j]<<endl;
}

void Punto::CalcularActiveClusters()
{
    numclusters=0;
    int i;
    for (i=0; i<Kmax; i++){
        CL[i].active=0;

        if (x[i]>0.5){
            CL[i].active = 1;
            numclusters++;
        }
    }
}

//Guardamos los valores de x que representan el centro del cluster i en CL[i].center[.]
void Punto::GuardarCenter(int cl)

```

```
{  
    int i;  
    for (i=0; i<columns; i++)  
        CL[cl].center[i]=x[Kmax + cl*columns + i];  
}
```

//Buscamos el cluster más cercano en función de las distancias de los puntos a los CL[i].center[..]

```
void Punto::AsignarPuntosACluster()  
{  
    int i,j,k;  
    double d,dmin;  
    int cl;  
  
    for (i=0; i<Kmax; i++)  
        CL[i].nk = 0;  
  
    for (j=0; j<rows; j++) {  
        asign[j] = 0;  
        dmin = 10000000000000000000;  
        for (i=0; i<Kmax; i++)  
        {  
            if (CL[i].active == 1){  
                d= SquareEuclideanDist(CL[i].center, matrix[j], columns);  
                if(d < dmin){  
                    dmin = d;  
                    cl=i;  
                }  
            }  
        }  
        asign[j] = cl;  
        CL[cl].nk++;  
    }
```

```
}
```

```
//Desactivamos los cluster vacíos
```

```
for (i=0; i<Kmax; i++)
    if ((CL[i].active==1) && (CL[i].nk == 0))
    {
        CL[i].active=0;
        numclusters--;
        if (numclusters==1) {
            // printf("(1) 1 solo cluster. Hit any key to continue> ");
            // getchar();
        }
    }
}
```

```
//Desactivamos los que tienen un solo elemento y lo asignamos a otro cluster
```

```
for (i=0; i<Kmax; i++)
    if(CL[i].nk == 1)
    {
        CL[i].active=0;
        CL[i].nk = 0;
        numclusters--;
        if (numclusters==1) {
            // printf("(2) 1 solo cluster. Hit any key to continue> ");
            // getchar();
        }
        for (j=0; j<rows; j++) {
            if(assign[j] == i){
                dmin = 1000000000000000000;
                for (k=0; k<Kmax; k++)
                {
                    if (CL[k].active==1){
                        d= SquareEuclideanDist(CL[k].center, matrix[j], columns);
```

```
        if(d < dmin){
            dmin = d;
            cl=k;
        }
    }
    asign[j] = cl;
    CL[cl].nk++;
}

}

for (i=0; i<Kmax; i++)
    if (CL[i].nk==1) {
        cout<<"Ojo! Tamaño demasiado pequeño"<<endl;
        printf("Hit any key to continue> ");
        getchar();
    }
}

double SquareEuclideanDist(double p1[dimcolumns],double p2[dimcolumns], int
lenght)
{
    double d=0;
    for (int i=0; i<lenght; i++)
        d+= (p1[i]-p2[i])*(p1[i]-p2[i]);

    return pow(d,0.5);
}

void Punto::CalcularCentroides()
```

```

{
    int i,j;
    for (i=0; i<Kmax; i++)
        //if (CL[i].active==1)
        for (j=0; j<columns; j++)
            CL[i].centroide[j]=0;

    for (i=0; i<rows; i++)
        for (j=0; j<columns; j++)
            CL[assign[i]].centroide[j] += matrix[i][j]/((double) CL[assign[i]].nk);
}

double Punto::CS() //Indicador de CS
{
    num = 0.0,den = 0.0;

    int i,j;
    for (i=0; i<Kmax; i++)
        for (j=0; j<Kmax; j++)
            between_dist[i][j]=0;

    for (i=0; i<Kmax; i++)
    {
        between_min[i] = 1E21;
        if ( (CL[i].active==1)&&(CL[i].nk>1))
        {
            int entra=0;
            for (j=0; j<Kmax; j++)
                if ((CL[j].active==1)&&(CL[j].nk>1))
                {
                    between_dist[i][j] = SquareEuclideanDist(CL[i].centroide, CL[j].centroide,
columns);

```



```
        if ((i!=j)&&(between_dist[i][j] < between_min[i])) {
            entra=1;
            between_min[i] = between_dist[i][j];
        }
    }
    if (entra==1)
        den += between_min[i];
}
}

for (j=0; j<rows; j++)
    within_max[j]=0;

for (i=0; i<Kmax; i++)
    within_mean[i]=0;

for (i=0; i<rows; i++)//Esta fila ya sabemos que pertenece al cluster asign[i]
    for (j=i+1; j<rows; j++)
    {
        if ( (i != j) && (asign[i]==asign[j]) && (distmatrix[i][j] > within_max[i]) )
            within_max[i] = distmatrix[i][j];
        if ( (i != j) && (asign[i]==asign[j]) && (distmatrix[i][j] > within_max[j]) )
            within_max[j] = distmatrix[i][j];
    }

for (i=0; i<rows; i++)
    if (CL[asign[i]].nk > 1)
        within_mean[asign[i]] += within_max[i]/CL[asign[i]].nk;
    else
        within_mean[asign[i]] = 0;

for (i=0; i<Kmax; i++)
```

```

        num += within_mean[i];

    return num/den;
}

double Punto::CH() //Indicador de Calinski-Harabasz
{
    num = 0.0, den = 0.0;

    int i,j;

    for (i=0; i<columns; i++)
        centro_datos[i]=0;

    for (i=0; i<rows; i++)
        for (j=0; j<columns; j++)
            centro_datos[j] += matrix[i][j]/rows;

    for (i=0; i<Kmax; i++)
        if ( (CL[i].active==1)&&(CL[i].nk>1))
            den += CL[i].nk * pow(SquareEuclideanDist(CL[i].centroide, centro_datos,
columns),1);

    for (i=0; i<rows; i++)//Esta fila ya sabemos que pertenece al cluster asign[i]
        num += pow(SquareEuclideanDist(CL[asign[i]].centroide, matrix[i], columns),1);

    return (num/den)*((double)(numclusters-1)/(rows-numclusters));
    //return -(den/num)*((double)(rows-numclusters)/(numclusters-1));
}

double Punto::DN()//Indicador de Dunn
{

```

```
num = -1E21,den = 1E21;

int i,j;
for (i=0; i<Kmax; i++)
    for (j=0; j<Kmax; j++)
        between_dist[i][j]=0;

for (i=0; i<Kmax; i++)
{
    //between_min[i] = 1E21;
    if ( (CL[i].active==1)&&(CL[i].nk>1))
    {
        int entra=0;
        for (j=0; j<Kmax; j++)
            if ((CL[j].active==1)&&(CL[j].nk>1))
            {
                between_dist[i][j] = SquareEuclideanDist(CL[i].centroide, CL[j].centroide,
columns);

                if ((i!=j)&&(between_dist[i][j] < den)) {
                    entra=1;
                    den = between_dist[i][j];
                }
            }
        //if (entra==1)
        // den += between_min[i];
    }
}

for (j=0; j<rows; j++)
    within_max[j]=0;
```

```

for (i=0; i<Kmax; i++)
    within_mean[i]=0;

for (i=0; i<rows; i++)//Esta fila ya sabemos que pertenece al cluster asign[i]
    for (j=i+1; j<rows; j++)
    {
        if ( (i != j) && (asign[i]==asign[j]) && (distmatrix[i][j] > within_max[i]) )
            within_max[i] = distmatrix[i][j];
        if ( (i != j) && (asign[i]==asign[j]) && (distmatrix[i][j] > within_max[j]) )
            within_max[j] = distmatrix[i][j];
    }

for (i=0; i<rows; i++)    //El diametro del cluster es la media de los máximos
    if (CL[asign[i]].nk > 1)
        within_mean[asign[i]] += within_max[i]/CL[asign[i]].nk;
    else
        within_mean[asign[i]] = 0;

for (i=0; i<Kmax; i++) //Guardamos en el numerador el máximo de todos los diámetros
medios
    if(within_mean[i] > num)
        num = within_mean[i];

return num/den;
}

double Punto::DB()//Indicador de Davies-Bouldin
{
    double indicator=0;

    int i,j;
    for (i=0; i<Kmax; i++)
        for (j=0; j<Kmax; j++)

```

```
        between_dist[i][j]=0;

for (i=0; i<Kmax; i++) //estos son los DL,m
    if ( (CL[i].active==1)&&(CL[i].nk>1))
        for (j=0; j<Kmax; j++)
            if ((CL[j].active==1)&&(CL[j].nk>1))
                between_dist[i][j] = SquareEuclideanDist(CL[i].centroide, CL[j].centroide,
columns);

for (j=0; j<rows; j++)
    within_max[j]=0;

for (i=0; i<Kmax; i++)
    within_mean[i]=0;

for (i=0; i<rows; i++)//Esta fila ya sabemos que pertenece al cluster asign[i]
    for (j=i+1; j<rows; j++)
    {
        if ( (i != j) && (asign[i]==asign[j]) && (distmatrix[i][j] > within_max[i]) )
            within_max[i] = distmatrix[i][j];
        if ( (i != j) && (asign[i]==asign[j]) && (distmatrix[i][j] > within_max[j]) )
            within_max[j] = distmatrix[i][j];
    }

for (i=0; i<rows; i++)
    if (CL[asign[i]].nk > 1)
        within_mean[asign[i]] += within_max[i]/CL[asign[i]].nk;
    else
        within_mean[asign[i]] = 0;

double DL[dimCluster];
```

```

for (i=0; i<Kmax; i++)
{
    if ( (CL[i].active==1)&&(CL[i].nk>1))
    {
        DL[i]=-1E21;

        for (j=i+1; j<Kmax; j++)
        {
            if ( (CL[j].active==1)&&(CL[j].nk>1))
            {
                double rate = (within_mean[i] + within_mean[j]) / between_dist[i][j];

                if (rate > DL[i]) {
                    DL[i] = rate;
                }
            }
        }
        if (DL[i]>=0)
            indicator += DL[i]/numclusters;
    }
}

return indicator;
}

double mean(double v[], int d)
{
    double suma=0;
    for (int i=0; i<d; i++) {
        suma += v[i];
    }
    return suma/((double)d);
}

```

```
double min(double v[], int d)
{
    double min=1000000;
    for (int i=0; i<d; i++) {
        if(v[i] < min)
        {
            min = v[i];
        }
    }
    return min;
}

double max(double v[], int d)
{
    double max=-100000;
    for (int i=0; i<d; i++) {
        if(v[i] > max)
            max=v[i];
    }
    return max;
}

double desvestp(double v[],int d)
{
    double suma=0, suma2=0;
    for (int i=0; i<d; i++) {
        suma += v[i];
        suma2 += v[i]*v[i];
    }
    return pow(suma2/((double)d)-pow(suma/((double)d),2),0.5);
}

void Punto::Ordenar_X()
```

```
{
    int i,j,k;
    Cluster temp;
    double aux;

    for (i=1; i<Kmax; i++) {
        for (j=0; j<Kmax-i; j++) {
            if (CL[j].nk > CL[j+1].nk) {
                temp =CL[j];
                CL[j] = CL[j+1];
                CL[j+1] = temp;

                //Y ahora odenamos las X
                aux = x[j];
                x[j] = x[j+1];
                x[j+1] = aux;

                for (k=0; k<columns; k++) {
                    aux = x[Kmax + j*columns + k];
                    x[Kmax + j*columns + k] = x[Kmax + (j+1)*columns + k];
                    x[Kmax + (j+1)*columns + k] = aux;
                }
            }
        }
    }
}
```